Balancing Competing Objectives for Welfare-Aware Machine Learning with Imperfect Data

Esther Rolf
UC Berkeley
esther_rolf@berkeley.edu

Max Simchowitz
UC Berkeley
msimchow@berkeley.edu

Sarah Dean
UC Berkeley
dean_sarah@berkeley.edu

Lydia T. Liu
UC Berkeley
lydiatliu@berkeley.edu

Daniel Björkegren
Brown University
danbjork@brown.edu

Moritz Hardt
UC Berkeley
hardt@berkeley.edu

Joshua Blumenstock
UC Berkeley
jblumenstock@berkeley.edu

Abstract

Although real-world decisions typically aim to balance many competing objectives, algorithmic decisions are often evaluated with a single objective function. This paper studies algorithmic policies which attempt to optimally trade off between two distinct measures of performance; i.e. profit and social utility, or user engagement and user health. While optimal policies can be described using traditional notions of Pareto optimality when high quality data are readily available, we focus on understanding how decisions should be made in noisy or data-poor regimes. We formalize this trade off and present empirical results on a real world dataset for content recommendation. These experiments underscore the applicability of our analyses and shed light on the nature of inherent trade offs in the application of machine learning methods to human-sensitive decisions.

1 Introduction

From financial loans [4] and humanitarian aid [5], to medical diagnosis [13] and criminal justice [3], consequential decisions in society increasingly rely on machine learning. In most cases, the machine learning algorithms used in these contexts are trained to optimize a single metric of performance. However, the decisions made by algorithms can have adverse side effects. Increasingly, the institutions which rely upon decision making algorithms are facing external and internal pressure to balance traditional private objectives (such as profit and user engagement) with public objectives that account for the well-being of those that they serve. In other words, most real-world decisions exist in a *multi-objective* setting, that requires the balance of multiple incentives and outcomes.

This paper develops a methodology for optimizing multi-objective decisions. Building on the traditional notion of Pareto optimality, which provides a concise characterization of optimal policies under complete information, we focus on understanding how to balance multiple objectives when those objectives are measured noisily or not directly observed. We believe this regime of imperfect information is far more common in real-world decisions, where one cannot easily measure the social consequences of an algorithmic decision. To show how the multi-objective framework can be used in practice, we present results using data from roughly 40,000 videos promoted by YouTube's recommendation algorithm. This illustrates the empirical trade-off between maximizing

AI for Social Good workshop at NeurIPS (2019), Vancouver, Canada.

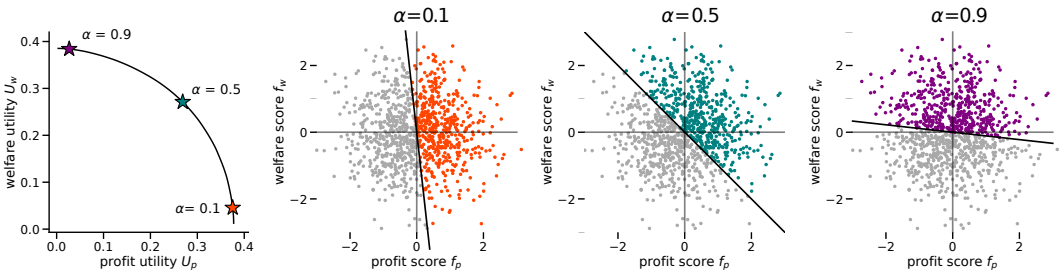


Figure 1: A Pareto curve (left) and the decision boundaries (right) induced by three different tradeoff parameters α. Colored points indicate selected individuals.

user engagement and promoting high-quality videos. We show that multi-objective optimization could produce substantial increases in average video quality at the expense of almost negligible reductions in user engagement.

Related Work Many notions of 'fairness' in machine learning have been proposed and studied [1, 2, 9]. The inherent trade-offs between fairness criteria are well-known [6, 19]. Previous works have also studied the tension between group-specific objectives and accuracy [18, 31], as well as highlighted the domain-specific nature of these concerns [12, 27]. An emerging line of work is concerned with the long-term impact of algorithmic decisions on societal welfare and fairness [10, 14, 26]. Previous investigation into the delayed impact of a decision policy on the wellbeing of different subpopulations [23] motivates us to study the direct optimization of welfare and profit, jointly. Complementary to previous work on multiobjective optimization [7, 8, 20] and learning Pareto frontiers [7, 17, 25], we are concerned with addressing uncertainty in the estimation of relevant objectives. We consider post-selection effects by focusing on a single round of algorithmic decisions, while some related work addresses the sequential nature of decisions on a longer term [22, 29]. For further exposition of multi-objective optimization in machine learning more broadly, we refer the reader to [15, 16].

2 Pareto-optimal policies

We consider a setting in which a policymaker has two simultaneous objectives: to maximize some private return (such as revenue or user engagement), which we generically refer to as *profit*; and to improve a public objective (such as social welfare or user health), which we refer to as *welfare*. The policymaker makes decisions about *individuals*, who are specified by feature vectors $x \in \mathbb{R}^d$.

Decision policies are functions that output a randomized decision $\pi(x) \in [0, 1]$ corresponding to the *probability* that an individual with features x is selected. We further associate to each individual a value p representing the expected profit to be garnered from approving an individual and w encoding the change in welfare. The profit and welfare objectives are thus:

$$\mathcal{U}_{\mathsf{W}}(\pi) = \mathbb{E}[w \cdot \pi(x)] \quad \text{and} \quad \mathcal{U}_{\mathsf{P}}(\pi) = \mathbb{E}[p \cdot \pi(x)]\,. \tag{1}$$

Given two objectives, one can no longer define a unique optimal policy π. Instead, we focus on policies π which are *Pareto-optimal* [28], in the sense that they are not strictly dominated by any alternative policy, i.e. there is no π' such that both profit and welfare objectives are strictly better under π'. Under general conditions, it is equivalent to consider policies that maximize a weighted combination of both objectives. We can thus parametrize the Pareto-optimal policies by $\alpha \in [0, 1]$:

$$\pi^\star_\alpha \in \operatorname{argmax} \mathcal{U}_\alpha(\pi), \quad \mathcal{U}_\alpha(\pi) := (1-\alpha)\mathcal{U}_{\mathsf{P}}(\pi) + \alpha\mathcal{U}_{\mathsf{W}}(\pi).$$

We defer a formal statement of this fact to the Appendix, in Proposition A.1.

Exact scores We first consider an idealized setting, where the welfare and profit contributions w and p can be directly determined from the features x via exact *score functions*, $f_{\mathsf{W}}(x) = w$, $f_{\mathsf{P}}(x) = p$. These exact score functions can be thought of as sufficient statistics for the decision: the expected weighted contribution from accepted individuals is described by $((1-\alpha)p + \alpha w)$.

Therefore, one can show that the optimal policy is given simply by thresholding this composite:

$$\pi^\star_\alpha(p, w) = \mathbb{I}((1-\alpha)p + \alpha w \geq 0). \tag{2}$$

Though they are all Pareto-optimal, the policies $\pi^\star_\alpha$ induce different trade-offs between the two objectives. The parameter α determines this trade-off, tracing the *Pareto frontier*.

$$\mathcal{P}_{\text{exact}} := \{(\mathcal{U}_{\mathsf{P}}(\pi^\star_\alpha), \mathcal{U}_{\mathsf{W}}(\pi^\star_\alpha)) : \alpha \in [0, 1]\}$$

Figure 1 plots an example of this curve (leftmost panel) and the corresponding decision rules for three points along it. We note the concave shape of this curve, a manifestation of *diminishing marginal returns*: as a decision policy forgoes profit to increase total welfare, less and less welfare is gained for the same amount of profit forgone.

Inexact scores In real applications, we typically do not know the profit score p or welfare score w — or even the score functions f_{P} and f_{W} — a priori. Instead, we might estimate score functions from data in the hope that these models can provide good predictions on future examples. Given access to finite samples $\{(x_i, w_i)\}_{i=1}^{n_w}$ and $\{(x_j, p_j)\}_{j=1}^{n_p}$ drawn from an underlying distribution, we can estimate the score functions via *empirical risk minimization*, where we set

$$\widehat{f}_{\mathsf{W}} \in \operatorname*{argmin}_f \frac{1}{n_{\mathsf{W}}} \sum_{i=1}^{n_{\mathsf{W}}} \ell_{\text{pred}}(f(x_i), w_i), \qquad \widehat{f}_{\mathsf{P}} \in \operatorname*{argmin}_f \frac{1}{n_{\mathsf{P}}} \sum_{i=1}^{n_{\mathsf{P}}} \ell_{\text{pred}}(f(x_j), p_j)\,.$$

Then we can define a selection rule based on α-defined *plug-in threshold policies* via

$$\pi^{\text{plug}}_\alpha(x) = \mathbb{I}((1-\alpha)\widehat{f}_{\mathsf{P}}(x) + \alpha\widehat{f}_{\mathsf{W}}(x) \geq 0)\,. \tag{3}$$

This policy is optimal when predicted scores exactly recover the truth (recovering the exact case (2)), and furthermore we can bound the sub-optimality gap when they do not (Appendix A.2). However, when the scores are imperfect, the plug-in policy may not be Pareto optimal.

Pareto-optimality for learned scores We now describe Pareto-optimal policies over Π_{emp}, the class of all policies that act on the predicted scores. First, we define the following conditional expectations over the distribution $\mathcal{D}$ of (x, p, w):

$$\overline{\mu}_{\mathsf{P}}(\widehat{f}_{\mathsf{P}}(x), \widehat{f}_{\mathsf{W}}(x)) := \mathbb{E}_{\mathcal{D}}[p \mid \widehat{f}_{\mathsf{P}}(x), \widehat{f}_{\mathsf{W}}(x)], \quad \overline{\mu}_{\mathsf{W}}(\widehat{f}_{\mathsf{P}}(x), \widehat{f}_{\mathsf{W}}(x)) := \mathbb{E}_{\mathcal{D}}[w \mid \widehat{f}_{\mathsf{P}}(x), \widehat{f}_{\mathsf{W}}(x)]\,.$$

Intuitively, these values represent our best guesses of p and w, given the predicted scores. We define π^{opt}_α as the threshold policy on the composite $(1-\alpha)\cdot\overline{\mu}_{\mathsf{P}} + \alpha\cdot\overline{\mu}_{\mathsf{W}}$.

Theorem 2.1 (Pareto Frontier in inexact knowledge case)**.** *The policies π^{opt}_α are Pareto-optimal over the class Π_{emp}. The associated empirical frontier $\mathcal{P}_{\text{emp}}$ is dominated by the exact frontier $\mathcal{P}_{\text{exact}}$.*

Thus an optimal empirical-score based policy can also be realized as a threshold policy (this time of the conditional expectations), and it obeys the same diminishing-returns phenomenon as in the exact score case. We present a proof of this result in Appendix A.3.

We emphasize that π^{opt}_α requires computing conditional expectations over the distribution $\mathcal{D}$, and therefore will differ from the plug in policy defined in (3). Nevertheless, π^{opt} and π^{plug} coincide as long as the predicted score functions are *well-calibrated*, in the sense that $\mathbb{E}[p \mid \widehat{f}_{\mathsf{P}}(x), \widehat{f}_{\mathsf{W}}(x)] = \widehat{f}_{\mathsf{P}}(x)$ and $\mathbb{E}[w \mid \widehat{f}_{\mathsf{P}}(x), \widehat{f}_{\mathsf{W}}(x)] = \widehat{f}_{\mathsf{W}}(x)$. One example of $\widehat{f}_{\mathsf{P}}(x)$ that achieves this is the conditional expectation of p given x, i.e., $\widehat{f}_{\mathsf{P}}(x) = \mathbb{E}[p \mid x]$. Calibration can be achieved by empirical risk minimization under typical conditions [24].

3 Empirical investigation: balancing user engagement and health

We now illustrate how the multi-objective framework can be used to balance the desire to promote high quality content with the need for profit. Specifically, we work with a dataset that contains measures of content quality and content engagement for 39,817 YouTube videos. These data were constructed as part of an independent effort to automatically ascertain the quality and truthfulness of YouTube videos [11].

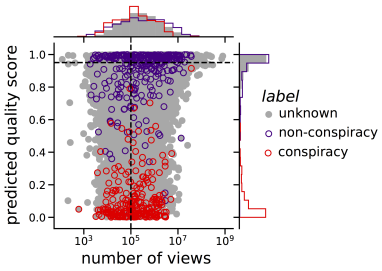
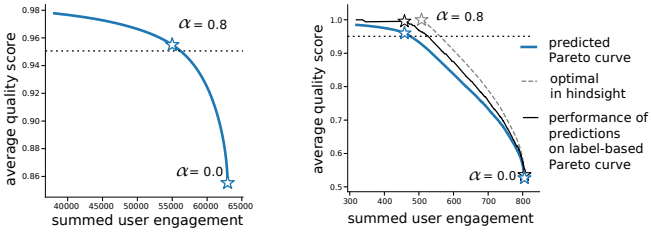


Figure 2: Distribution of YouTube data predicted quality scores across unlabled videos (grey), and hand labeled conspiracy (red) and non-conspiracy (purple) videos.



(a) (b)

Figure 3: (a) Estimated Pareto curve for YouTube predictions. Stars indicate specific α trade offs. (b) Estimated Pareto curve for YouTube predictions on labeled data subset (thick blue line). Optimal-in-hindsight curve (dashed grey line) and performance of predictions on label set (black line).

The measure of quality $\widehat{f}_{\mathsf{W}}$ we use is a function of the 'conspiracy score' developed by [11], which indicates the probability that the video promotes a debunked conspiracy theory. From this predicted score $s_{\text{conspiracy}}$ we derive a predicted 'quality score' as $(1 - s_{\text{conspiracy}}) \in [0, 1]$. Defining an allowable quality threshold as the median score of all videos ($= 0.95$), we instantiate $\widehat{f}_{\mathsf{W}} = (1 - s_{\text{conspiracy}}) - 0.95$. It is worth noting that no notions of engagement (e.g. view count, comment count) were included as training data to learn $s_{\text{conspiracy}}$.

We instantiate the engagement (profit) score $f_{\mathsf{P}}[i]$ for video i as $\log((1 + \#\,\text{views}[i])/100,000)$, where the number of views is observed directly. Dividing by a large constant represents that videos with low view counts may not be profitable due to storage and hosting costs, i.e. videos with view counts below $100,000$ (roughly 32% of the videos in the validation set) do not break a profit margin. The resulting distribution over f_{P} and $\widehat{f}_{\mathsf{W}}$ is shown in Figure 2 (grey dots), where the thresholds in each score are denoted with dotted lines.

Using these scores and predictions, we estimate a Pareto frontier using the optimal policies $\pi_{\alpha}^{\text{plug}}$ for learned scores from Eq. (3). The resulting estimated Pareto curve is shown in Figure 3a.

The curve is concave, demonstrating the phenomenon of diminishing returns in the trade off between total user engagement and average video quality. While there is always some quality to gain by sacrificing some total engagement, these relative gains are greatest when the starting point is close to an engagement-maximizing policy. Specifically, at the maximum-engagement end of the spectrum (lower right star), we can gain a 1.0% increase in average video quality for a 0.1% loss in total engagement. However, for a policy that already with trade off rate $\alpha = 0.8$ (upper left star), to obtain an increase of 0.9% in welfare, a larger loss of 6.5% in user engagement is required.

Next, we assess the validity of this estimated Pareto curve using the small set of hand-labeled training set instances from which $s_{\text{conspiracy}}$ was learned. This set consists of 541 video instances which are hand-labeled as either conspiracy (251) or non-conspiracy (290), as well as their view counts and predictions. As shown in Figure 2, these validation points are drawn from a different distribution, and thus tend to lie toward the extremes of the quality measure. This assessment is likeley optimistic due to the fact that the scores predictor functions were trained on this same data; nonetheless, this is an important check to perform on the estimated Pareto frontier.

In Figure 3b we plot the optimal-in-hindsight Pareto frontier (dashed grey line) if we had known the labels a priori and applied thresholds according to (2). We also plot the performance of our estimated policy $\pi_{\alpha}^{\text{plug}}$ on the labeled instances (black line). The stars on each curve correspond to decision thresholds with $\alpha = 0$ and $\alpha = 0.8$, and illustrate the calibration of the curves.

Relating back to Theorem 2.1, we see that performance of the learned scores (black line) is dominated by that of the optimal classifier, as is the predicted Pareto curve (thick blue line). We note that the predicted Pareto curve under-predicts the actual performance; in general it is possible for the opposite to be true. Importantly, we observe that the curves representing the predicted and actual performance show similar qualitative trade offs.

4 Conclusion

We have presented a methodology for developing welfare-aware policies that jointly optimize private institutional objectives with public objectives involving social welfare. Taking care to consider data-poor regimes, we develop theory around the optimality of using learned predictors to make decisions. Our experiments corroborate our theoretical result, showing that thresholding on predicted scores can approach a Pareto optimal policy. Ongoing and future work is focused on extending this framework to approach more intricate tradeoffs, such as the cost-benefit trade-off between spending excess profit to collect more data versus traversing a sub-optimal Pareto frontier. Altogether, this work provides encouraging insight into how to address the trade-offs inherent to designing machine-learning based welfare aware policies while emphasizing data collection and measurement as a crucial component.

Acknowledgement

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1752814, the Bill and Melinda Gates Foundation, and the Center for Effective Global Action (Digital Credit Observatory).

References

[1] Solon Barocas and Andrew D Selbst. Big data's disparate impact. *UCLA Law Review*, 2016.

[2] Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2018. http://www.fairmlbook.org.

[3] Richard Berk. *Criminal Justice Forecasts of Risk: A Machine Learning Approach*. Springer Science & Business Media, April 2012. ISBN 978-1-4614-3085-8. Google-Books-ID: Jrlb6Or8YisC.

[4] Daniel Björkegren and Darrell Grissen. Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment. *arXiv:1712.05840 [cs]*, December 2017. URL http://arxiv.org/abs/1712.05840. arXiv: 1712.05840.

[5] Joshua Evan Blumenstock, Gabriel Cadamuro, and Robert On. Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076, November 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aac4420. URL http://www.sciencemag.org/content/350/6264/1073.

[6] A. Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5, 2017.

[7] Kalyanmoy Deb and Deb Kalyanmoy. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 2001. ISBN 047187339X.

[8] Jean-Antoine Désidéri. Multiple-gradient descent algorithm (mgda) for multiobjective optimization. *Comptes Rendus Mathematique*, 350(5-6):313–318, 2012.

[9] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, pages 214–226, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1115-1. doi: 10.1145/2090236.2090255. URL http://doi.acm.org/10.1145/2090236.2090255.

[10] Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. Runaway Feedback Loops in Predictive Policing. *arXiv:1706.09847 [cs, stat]*, June 2017. URL http://arxiv.org/abs/1706.09847. arXiv: 1706.09847.

[11] Marc Faddoul, Guillame Chaslot, and Hany Farid. A longitudinal analysis of youtube's promotion of conspiracy videos. *In Preparation*, 2019.

[12] Marc Fleurbaey and Francois Maniquet. Optimal income taxation theory and principles of fairness. *Journal of Economic Literature*, 56(3):1029–79, 2018.

[13] Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros,

Ramasamy Kim, Rajiv Raman, Philip C. Nelson, Jessica L. Mega, and Dale R. Webster. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, 316(22):2402–2410, December 2016. ISSN 0098-7484. doi: 10.1001/jama.2016.17216. URL `https://jamanetwork.com/journals/jama/fullarticle/2588763`.

[14] Lily Hu and Yiling Chen. Welfare and Distributional Impacts of Fair Classification. *arXiv:1807.01134 [cs, stat]*, July 2018. URL `http://arxiv.org/abs/1807.01134`. arXiv: 1807.01134.

[15] Yaochu Jin. *Multi-objective machine learning*, volume 16. Springer Science & Business Media, 2006.

[16] Yaochu Jin and Bernhard Sendhoff. Pareto-based multiobjective machine learning: An overview and case studies. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(3):397–415, 2008.

[17] Il Yong Kim and Oliver L de Weck. Adaptive weighted-sum method for bi-objective optimization: Pareto front generation. *Structural and multidisciplinary optimization*, 29(2):149–158, 2005.

[18] Michael P Kim, Amirata Ghorbani, and James Zou. Multiaccuracy: Black-box post-processing for fairness in classification. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 247–254. ACM, 2019.

[19] Jon M. Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. *Proc. 8th ITCS*, 2017.

[20] Joshua Knowles. Parego: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 10(1): 50–66, 2006.

[21] Hidetoshi Komiya. Elementary proof for sion's minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7, 1988.

[22] Chunming Liu, Xin Xu, and Dewen Hu. Multiobjective reinforcement learning: A comprehensive overview. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(3):385–398, 2014.

[23] Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. Delayed impact of fair machine learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3156–3164, Stockholm, Sweden, 2018.

[24] Lydia T. Liu, Max Simchowitz, and Moritz Hardt. The implicit fairness criterion of unconstrained learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4051–4060, Long Beach, California, USA, 2019. PMLR.

[25] Ilya Loshchilov, Marc Schoenauer, and Michèle Sebag. A mono surrogate for multiobjective optimization. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, pages 471–478. ACM, 2010.

[26] Hussein Mouzannar, Mesrob I. Ohannessian, and Nathan Srebro. From Fair Decision Making to Social Equality. *arXiv:1812.02952 [cs, stat]*, December 2018. URL `http://arxiv.org/abs/1812.02952`. arXiv: 1812.02952.

[27] Alejandro Noriega, Bernardo Garcia-Bulle, Luis Tejerina, and Alex Pentland. Algorithmic fairness and efficiency in targeting social welfare programs at scale. *Bloomberg Data for Good Exchange Conference*, 2018.

[28] Vilfredo Pareto. *Manuale di economia politica*, volume 13. Societa Editrice, 1906.

[29] Kristof Van Moffaert and Ann Nowé. Multi-objective reinforcement learning using sets of pareto dominating policies. *The Journal of Machine Learning Research*, 15(1):3483–3512, 2014.

[30] Larry Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.

[31] Indre Zliobaite. On the relation between accuracy and fairness in binary classification. *arXiv preprint arXiv:1505.05723*, 2015.

A Proofs for General Characterization of Pareto Curves

A.1 Pareto Policies Optimize Weighted Combination of Utilities

Proposition A.1 (Pareto optimal policies optimize a composite objective)**.** *If and only if a policy* $\pi^\star \in \Pi$ *is Pareto optimal, there exists an* $\alpha \in [0,1]$ *for which*

$$\pi^\star \in \operatorname*{argmax}_{\pi\in\Pi} \mathcal{U}_\alpha(\pi)$$
$$\mathcal{U}_\alpha(\pi) := (1-\alpha)\mathcal{U}_\mathsf{P}(\pi) + \alpha\mathcal{U}_\mathsf{W}(\pi).$$

Proof. First, we prove that if $\pi^\star \in \operatorname{argmax}_\pi \mathcal{U}_\alpha(\pi) := (1-\alpha)\mathcal{U}_\mathsf{P}(\pi) + \alpha\mathcal{U}_\mathsf{W}(\pi)$, then $\pi^\star$ is Pareto optimal. Suppose that there exists an α for which $\pi^\star \in \operatorname{argmax}_\pi \mathcal{U}_\alpha(\pi)$. If $\alpha \in \{0,1\}$, then $\pi^\star$ maximizes either $\mathcal{U}_\mathsf{W}(\cdot)$ or $\mathcal{U}_\mathsf{P}(\cdot)$, and is therefore Pareto optimal by definition. Otherwise, if $\alpha \in (0,1)$, suppose for the sake of contradiction that $\pi^\star$ is not Pareto optimal. Then exists a policy π for which either $\mathcal{U}_\mathsf{W}(\pi^\star) \le \mathcal{U}_\mathsf{W}(\pi)$ and $\mathcal{U}_\mathsf{P}(\pi^\star) \le \mathcal{U}_\mathsf{P}(\pi)$, and that one of these inequalities is strict. We can then check that $\mathcal{U}_\alpha(\pi^\star) < \mathcal{U}_\alpha(\pi)$, contradiction that $\pi^\star \in \operatorname{argmax}_\pi \mathcal{U}_\alpha(\pi)$.

To show the other direction, suppose that $\pi^\star$ is Pareto optimal. If $\pi^\star$ maximizes either profit or welfare, then $\pi^\star \in \operatorname{argmax}_\pi \mathcal{U}_\alpha$ for either $\alpha = 1$ or $\alpha = 0$. Otherwise, let $W = \mathcal{U}_\mathsf{W}(\pi^\star)$. Then, by Pareto optimality,

$$\begin{aligned}
\pi^\star &\in \operatorname{argmax}\{\mathcal{U}_\mathsf{P}(\pi) : \mathcal{U}_\mathsf{W}(\pi) \ge \mathcal{U}_\mathsf{W}(\pi^\star), \pi \in \Pi\} \\
&= \operatorname*{argmax}_{\pi\in\Pi} \left(\mathcal{U}_\mathsf{P}(\pi) + \min_{t\ge 0} t(\mathcal{U}_\mathsf{W}(\pi) - \mathcal{U}_\mathsf{W}(\pi^\star) \right) \\
&= \operatorname*{argmax}_{\pi\in\Pi} \min_{t\ge 0} \left(\mathcal{U}_\mathsf{P}(\pi) + t\left(\mathcal{U}_\mathsf{W}(\pi) - \mathcal{U}_\mathsf{W}(\pi^\star)\right)\right).
\end{aligned}$$

The map $\mathcal{U}_\mathsf{P}(\pi)$ and $\mathcal{U}_\mathsf{W}(\pi)$ are both linear functions in π. Hence, if Π is a a convex, and compact in a topology in which $\pi \mapsto \mathcal{U}_\mathsf{P}(\pi)$ and $\mathcal{U}_\mathsf{W}(\pi)$ are continuous, Sion's minimax theorem [21] ensures that strong duality holds, which means that we can switch order of the minimization over t and maximization over π. Thus, for some $t \ge 0$,

$$\begin{aligned}
\pi^\star \in \operatorname*{argmax}_\pi \left(\mathcal{U}_\mathsf{P}(\pi) + t\cdot\mathcal{U}_\mathsf{W}(\pi) - t\cdot\mathcal{U}_\mathsf{W}(\pi^\star)\right) &= \operatorname*{argmax}_\pi \left(\mathcal{U}_\mathsf{P}(\pi) + t\cdot\mathcal{U}_\mathsf{W}(\pi)\right) \\
= \operatorname*{argmax}_\pi \left(\frac{1}{1+t}\mathcal{U}(\pi) + \frac{t}{1+t}\mathcal{U}_\mathsf{W}(\pi)\right) &= \operatorname*{argmax}_\pi \left(\mathcal{U}_{1/(1+t)}(\pi)\right),
\end{aligned}$$

as needed. The convexity of Π means that Π is closed under the randomized combination of policies. In the simplest case, compactness is achieved when the space of features is finite (e.g. features x can only take a values in a discrete, finite subset of $\mathbb{R}^d$). □

A.2 Utilitity Loss induced by Score Function Suboptimality

Proposition A.2 (Sub-optimality bound)**.** *The gap in* α*-utility from applying policy* (3) *versus applying the optimal policy* (2) *is bounded as*

$$\mathcal{U}_\alpha(\pi^\star_\alpha) - \mathcal{U}_\alpha(\pi^{\text{plug}}_\alpha) \le (1-\alpha)\mathbb{E}[|\widehat{f}_\mathsf{P}(x) - f_\mathsf{P}(x)|] + \alpha\mathbb{E}[|\widehat{f}_\mathsf{W}(x) - f_\mathsf{W}(x)|]. \tag{4}$$

In statistical learning settings without model misspecification, $|\widehat{f}_\mathsf{P}(x) - f_\mathsf{P}(x)|$ typically decreases at the rate $\frac{1}{\sqrt{n_\mathsf{P}}}$ [30]. In practice, we can estimate $\mathbb{E}[|\widehat{f}_\mathsf{P}(x) - f_\mathsf{P}(x)|]$ with the empirical absolute error on a validation set (and similarly for w), and thus estimate the upper bound empirically.

Proof of Proposition A.2. We compute

$$\mathcal{U}_\alpha(\pi^{\text{plug}}_\alpha) - \mathcal{U}_\alpha(\pi_\alpha) = \mathbb{E}[((1-\alpha)p + \alpha w)\left(\pi^{\text{plug}}_\alpha - \pi_\alpha\right)]$$

Define the functions $Y(x) = (1-\alpha)f_\mathsf{P}(x) + \alpha f_\mathsf{W}(x)$, and let $E(x) = (1-\alpha)(\widehat{f}_\mathsf{P}(x) - f_\mathsf{P}(x)) + \alpha(\widehat{f}_\mathsf{W}(x) - f_\mathsf{W}(x))$. Then, $\pi^{\text{plug}}_\alpha(x) - \pi_\alpha = \mathbb{I}(Y(x) + E(x) \ge 0) - \mathbb{I}(Y(x) \ge 0)$. We see that this difference is at most 1 in magnitude, and is 0 unless possibly if $|Y(x)| \le |E(x)|$. Hence,

$$|Y(x)| \cdot |\pi^{\text{plug}}_\alpha(x) - \pi_\alpha(x)| \le |E(x)|$$

Therefore

$$\begin{aligned}
|\mathcal{U}_\alpha(\pi_\alpha^{\text{plug}}) - \mathcal{U}_\alpha(\pi_\alpha)| &= |\mathbb{E}[Y(x)(\pi_\alpha^{\text{plug}}(x) - \pi_\alpha(x)]| \\
&\leq \mathbb{E}[|Y(x)| \cdot |\pi_\alpha^{\text{plug}}(x) - \pi_\alpha(x)|] \\
&\leq \mathbb{E}[|E(x)|] = \mathbb{E}[|(1-\alpha)(\widehat{f}_{\mathsf{P}}(x) - f_{\mathsf{P}}(x)) + \alpha(\widehat{f}_{\mathsf{W}}(x) - f_{\mathsf{W}}(x))|] \\
&\leq (1-\alpha)\mathbb{E}[|\widehat{f}_{\mathsf{P}}(x) - f_{\mathsf{P}}(x)|] + \alpha\mathbb{E}[|\widehat{f}_{\mathsf{W}}(x) - f_{\mathsf{W}}(x))|].
\end{aligned}$$

□

A.3 Proof of Theorem 2.1

We first show that

$$\pi_\alpha^{\text{opt}} \in \underset{\pi \in \Pi_{\text{emp}}}{\operatorname{argmax}}\, \mathbb{E}\left[\mathcal{U}_\alpha(\pi(\widehat{f}_{\mathsf{W}}, \widehat{f}_{\mathsf{P}}))\right].$$

We have that

$$\begin{aligned}
\mathcal{U}_\alpha(\pi) &= \mathbb{E}[((1-\alpha)p + \alpha w))\, \pi(\widehat{f}_{\mathsf{P}}, \widehat{f}_{\mathsf{W}})] \\
&= \mathbb{E}[\Big((1-\alpha)\mathbb{E}[p \mid \widehat{f}_{\mathsf{P}}, \widehat{f}_{\mathsf{W}}] + \alpha\mathbb{E}[w \mid \widehat{f}_{\mathsf{P}}, \widehat{f}_{\mathsf{W}}]\Big) \cdot \pi(\widehat{f}_{\mathsf{P}}, \widehat{f}_{\mathsf{W}})] \\
&:= \mathbb{E}[\Big((1-\alpha)\overline{\mu}_{\mathsf{P}}(\widehat{f}_{\mathsf{P}}, \widehat{f}_{\mathsf{W}}) + \alpha\overline{\mu}_{\mathsf{W}}(\widehat{f}_{\mathsf{P}}, \widehat{f}_{\mathsf{W}})\Big) \cdot \pi(\widehat{f}_{\mathsf{P}}, \widehat{f}_{\mathsf{W}})] \\
&\leq \mathbb{E}\left[\max\left\{(1-\alpha)\overline{\mu}_{\mathsf{P}}((\widehat{f}_{\mathsf{P}}, \widehat{f}_{\mathsf{W}}) + \alpha\overline{\mu}_{\mathsf{W}}(\widehat{f}_{\mathsf{P}}, \widehat{f}_{\mathsf{W}}), 0\right\}\right] \quad = \mathcal{U}_\alpha(\pi_\alpha^{\text{opt}}).
\end{aligned}$$

Hence, we obtain the Pareto optimality of π_α^{opt} by Proposition A.1.

Moreover, empirical policies are dominated by those induced by the true score functions because, as established, the Pareto optimal policies based on the true score functions are in fact Pareto optimal over all policies that are induced by a function of the features x.