

Fairness and classification results on Credit Data Set

June 2019

1 Measurement

- Statistical Parity (SP) [1]: $SP(f) = |p(f(x) = 1|s = 1) - p(f(x) = 1|s = 0)|$.
- Normed Disparate (ND) [1]: $ND(f) = \left| \frac{p(f(x)=1|s=1)}{p(f(x)=1|s=0)} - 1 \right|$.
- Classifier Error: it is the standard classification error.
- Error Parity: $EP(f) = |er(f|s = 1) - er(f|s = 0)|$.
- Error Disparate: $ED(f) = \left| \frac{er(f|s=1)}{er(f|s=0)} - 1 \right|$.

2 Experimental results

Table 1: Classification Performance on the Credit Data Set

Method	Statistical Parity	Normed Disparate	Classifier Error	Error Parity	Error Disparate
FRR [2]	.0186±.0016	.2914±.0186	.2296±.0058	.0302	.1184
FKRR [3]	.0079±.0011	.1170±.0117	.2001±.0054	.0374	.1643
FGR [4]	.0779±.0571	.1283±.0987	.2412±.0469	.0253	.0951
FPCA1 [5]	.1716±.0149	.1458±.0234	.4025±.0382	.0941	.2277
FPCA2 [6]	.0981±.0164	.1048±.0193	.3224±.0045	.0663	.1859
LFR [7]	.1534±.0132	.2011±.0133	.2772±.0051	.0471	.1663

References

- [1] D. McNamara, C. S. Ong, and R. C. Williamson, “Provably fair representations,” *CoRR*, 2017.
- [2] T. Calders, A. Karim, F. Kamiran, W. Ali, and X. Zhang, “Controlling attribute effect in linear regression,” in *2013 IEEE 13th International Conference on Data Mining*. IEEE, 2013, pp. 71–80.
- [3] A. Pérez-Suay, V. Laparra, G. Mateo-García, J. Muñoz-Marí, L. Gómez-Chova, and G. Camps-Valls, “Fair kernel learning,” in *ECMLPKDD*, 2017.
- [4] T. Kamishima, S. Akaho, and J. Sakuma, “Fairness-aware learning through regularization approach,” in *ICDM Workshops*, 2011.
- [5] S. Samadi, U. Tantipongpipat, J. H. Morgenstern, M. Singh, and S. Vempala, “The price of fair pca: One extra dimension,” in *NIPS*, 2018.
- [6] M. Olfat and A. Aswani, “Convex formulations for fair principal component analysis,” *CoRR*, 2018.
- [7] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork, “Learning fair representations,” in *ICML*, 2013.