Research paper

Author’s name

Author’s affiliation

City & country location

E-mail address.

Abstract - We look at whether PCA's typical dimensionality reduction approach provides data representations with varied fidelity for two distinct populations by accident. We demonstrate that PCA has a bigger reconstruction error on population A than on population B in numerous real-world data sets (for example, women versus men or lower- versus higher-educated individuals). Even if the data set includes a same number of samples from A and B, this may happen. This drives our research into dimensionality reduction strategies that retain integrity for both A and B. We define Fair PCA and describe a polynomial-time technique for locating a low-dimensional data representation that is approximately optimum in terms of this measure. Finally, we demonstrate that our technique can be utilized to efficiently produce a fair low-dimensional representation of data using real-world data sets.

INTRODUCTION

Concerns about the fairness of rules and judgments based on protected groups (such as race or gender) are rising, notwithstanding the effectiveness of machine learning in guiding policies and automating decision making (Miller 2015; Rudin 2013; Angwin et al. 2016; Munoz, Smith, and Patil 2016). That's why many researchers are focusing on defining fairness in supervised learning and developing learners that preserve high prediction accuracy while also lowering injustice; Calders, Kamiran and (Berk et al. 2017).

First and foremost, unsupervised learning is often used as a pre-processing step for other learning approaches. It's possible that fair dimensionality reduction, which is often done prior to clustering, might give ways for fair clustering. Fair unsupervised learning approaches may be integrated with state-of-the-art supervised learning techniques to create new fair supervised learners. Though this paper's focus is on techniques that have been developed to generate fair and accurate data transformations that maintain high accuracy for classifiers that make predictions using transformed data, these previous studies are correctly categorised as "supervised learning" since the data transformations are calculated with respect to the label used for predictions (Dwork et al. 2012; Zemel and Feldman, 2013; Feldman and Zemel, 2015).

This is a quick overview of the subject matter. Individuals may be mapped to probability distributions across potential classes using a linear program proposed by Dwork et al. (2012). In a non-convex formulation, Calmon et al. (2017) provide an intermediate representation for fair clustering. For matching feature distributions based on the protected property, Feldman et al. (2015) present an approach for scaling data points.

BACKGROUND

Outlines and new ideas are welcome.

In this study, we focus on the fairness of principal component analysis (PCA). A new quantitative concept of fairness for dimensionality reduction is proposed and shown in Section 3. Convex maximization formulae for equitable PCA as well as kernel PCA are developed in Section 5. Our semidefinite programming (SDP) formulations are shown in Section 6 employing a wide range of data, including fair PCA as a pre-processing step for fair (within the context of age) grouping of health information that might influence health assurance prices.

Notation

Heaviside function and the vector e are defined as [n] = [1,...,n], 1(u), and [n] = [1,...,n]. Negative semidefinite matrices with q-q dimensions may be described as U ∈ Sq+ (or when dimensions are

clear). The inner product is denoted by h., I, while the identity matrix is denoted by I.

Data consists of 2-tuples (xi, zi) for each of the integers 1 through n. Each pair of 2-tuples (xi,zi) represents a collection of characteristics, and each zi identifies a protected class (where xi = 1,...,n). When dealing with a matrix W, the i-th row of W is symbolized by the symbol Wi. Suppose that X ∈ Rn x p and Z ∈ Rn are the matrices such that (xi x) T is equal to (xi x) T and Zi is equal to zi.

Let K (X, X0) = [k(Xi, Xj0)]ij be the converted Gram matrix for a kernel function k: Rp x Rp → R+, where Rp x Rp R+ is the radial kernel function. As part of applying the kernel trick, we must replace xTi xj with K(xi,xj).

Fairness for dimensionality reduction

In supervised learning, definition Specification of the fact that predictions conditioned on the protected class are about equal. However, since predictions are not generated in unsupervised learning, similar fairness principles cannot be used to dimensionality reduction. This section examines the concept of fairness in the context of dimensionality reduction. A broad quantitative definition of fairness is first presented and justified, and then various key examples in which this concept is used are discussed.

We argue that a dimensionality reduction $\Pi : R^p \rightarrow R^d$ is $\Delta(h)$-fair if

$$\Big|\mathbb{P}\big[h(\Pi(x),t)=+1\big|z=+1\big] - \mathbb{P}\big[h(\Pi(x),t)=+1\big|z=-1\big]\Big| \leq \Delta(h),\ \forall t \in \mathbb{R}. \quad (1)$$

For classifiers, equation one is equivalent to detach effect, where we need that intervention should not differ at all across protected groups. This has been critiqued for being excessively stringent in categorization, and therefore other definitions of fairness, like similar odds and equalized chances, have been established. The emphasis here is on reducing mistake rates rather than ensuring that everyone is treated equally, as would be the case in lending, where equal chances would need treating all applicants with identical FICO scores equally, but differential effect would necessitate treating all applicants equally. In circumstances when y and z are significantly connected, this may be recommended. We can simply condition the variables on the opposite side of Equation one on the primary denotation, y, and include it into our model

Motivation

The concept given above is relevant for proportionality reduction since it suggests that a superintend learner employing fair proportionality-decreased data will be fair in and of itself. This is formalized in the next section:

Proposition 1. *Let's assume that we have a class of classifiers F, and that we can reduce the dimensions of this family by a factor of (F). To forecast a label y > 1,+1, any classifier picked from F to do so will have an influence on disparity that is smaller than (F).*

Prop. one is a logical extension of our notion of fairness. Most of the time, the purpose of the proportionality cutting is not to forecast the safeguarded groups. To prevent discrimination from occurring by accident while classifying with the family F or drawing qualitative conclusions from the outcomes of unsupervised learning, we used a conservative limit on intentional discrimination on z.

Particular cases

An essential particular case of our denotation is shown for the class $F_c = \{h(u,t) = \mathbf{1}(u \leq w + t) : w \in R^d\}$,

element-by-element interpretation of the inequality in this statement We can rewrite our definition in this scenario.

as $\sup_u \left|F_{\Pi(x)|z=+1}(u) - F_{\Pi(x)|z=-1}(u)\right| \leq \Delta(\mathcal{F}_c)$, where F is the C.D.F of the random variable for the denotation and R denotes the random variable in the superscript. If we express our definition in terms of a constraint on the Kolmogorov distance between two points Π (x) under the assumption that z = ± 1 is true, we get the following definition for this family.

Next, we precisely describe factual approximation of Δ(F). An factual approximate of Δ(*h*) is given by

$\widehat{\Delta}(h) = \sup_t |\frac{1}{\#P}\sum_{i\in P}\mathbf{1}(h(\Pi(x),t) = +1) - \frac{1}{\#N}\sum_{i\in N}\mathbf{1}(h(\Pi(x),t) = +1)|$. in the same way, we define

Δ(F) = sup{Δ(*h*) | *h* ∈ F}.

Proposition 2. *Suppose we have a fixed set of categorises F and the samples (xi,zi) are all identical, then any > 0 has a probability of at least one.* $- e\exp(-n\delta^2/2)$ *that* Δ(F) ≤

Δ(b F) + 8^pV(F)/*n* + *δ,*

The triangle inequality restricts (F) to (F) plus a stereotype error, for which Dudley's entropy integral provides conventional bounds, which is a generalization error plus a generalization error (Wainwright 2017).

First and foremost, remark 1. Remember that V(Fc) = d+1, and that V(Fv) = d + 1 (Shorack and Wellner 2009).

(Wainwright 2017). This indicates that when n is big in comparison to the final output, and Δ (b Fv) will be correct.

RELATED WORK

Fair grouping of health data

Health assurance firms are exploring using activity tracker data to change premiums for people based on their physical activity habits. An examination of new clustering data shows that different types of physical exercise are linked to various health outcomes (Fukuoka et al. 2018). An insurer's goal in looking for qualitative patterns in a customer's physical activity is to assist them classify the risks that they represent. In other words, since people who participate in the same types of activities tend to have comparable levels of medical expenses, it would be advantageous to design features that make it easier for insurers to group their clients together. However, there are a variety of legal fairness issues for health insurance prices when it comes to gender, ethnicity, and age. Therefore, if the patterns utilized to change prices have an excessively unfavourable effect on a given gender, race, or age, an insurance firm might be held accountable. If an insurer is looking to avoid discrimination on the basis of protected characteristics, feature engineering may be of interest. As a result of this, we use FPCA to evenly distribute our physical activity throughout the day. In our research, we are looking for patterns in actions that are predictive of an people's action patterns and consequently health concerns, but that are fair with regard to their age and gender. We apply immediate data obtained from the National Health and

Nutrition Examination Survey from 2005–2006

(The Centres for Disease Control and Prevention (CDCP) is a non-profit organization dedicated to preventing (CDC). The National Centre for Health Statistics (NCHS) published a study in 2018 that looked at the levels of physical activity of around 6000 women.

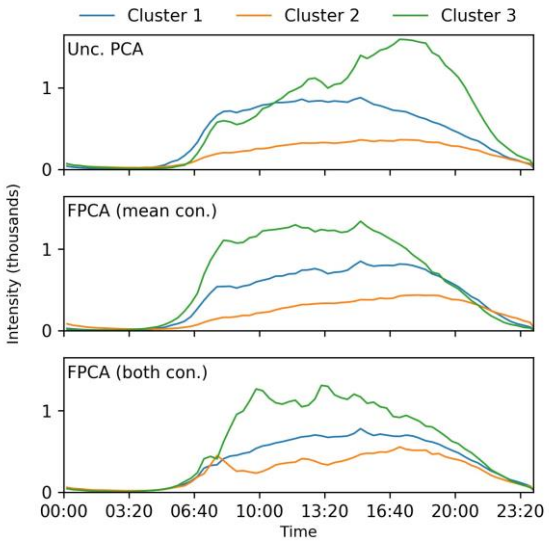

Figure 3: It is possible to depict the mean intensity of physical activity for each cluster created following proportionality reduction using PCA, FPCA with the mean constraint or FPCA with both constraints. One cluster's average activity level is shown by a single line in each graph.

In this case, the protected variable is a person's age, precisely whether or not they are above 40. During the weekdays, we average the activity data for each person into 20-minute buckets, excluding weekends from our study. As a result, for every participant, we have data detailing her average daily action level. After averaging, we omit those under the age of 12 and those who have spent more than 16 hours inactive in a single day. The top 1% of participants, as well as corrupted data, were omitted from the study. Data points that have been corrupted or inaccurate owing to accelerometer malfunction were omitted from this study. For this pre-processing, insurers and the patchiness of accelerometer data are taken into consideration, as is Fukuoka et al.

Prior to clustering, PCA may be employed as a pre-processing step to speed up the process. In this spirit, we use PCA, FPCA with the mean constraint, and FPCA with both restrictions, with = 0 and 0.1 throughout to determine the top five main components. Once the data has been reduced in dimension, we do k-means clustering (k = 3) on each example. Clusters are

shown in Figure 3 with the averaging of physical activity patterns for each instance. The percentage of examinees above the age of 40 in each cluster is also shown in Table 2. An policyholder is concerned in detailing an people's risk could focus mostly on their night-time activities under an unconstrained PCA, since the clusters are most distinct after some time. As seen in Table 2, there is a noticeable age differential between buckets when this strategy is used. The percentage of people above the age of 40 in each cluster. Almost three-quarters of those polled are above the age of 40. The standard deviation of the first three values is shown in the fourth row. As a realistic measure of fairness, it would be best if all clusters had the same age composition.

	UNC.	MEAN	BOTH
Cluster 1	43.18%	33.54%	35.61%
Cluster 2	32.94%	38.64%	36.11%
Cluster 3	8.71%	33.32%	37.28%
Std. Dev	14.87%	2.46%	1.79%

danger of unlawful pricing discrimination by the insurer Clustering clients grounded on their actions throughout the day, between 8:00 and 5:00 PM, seems to be less susceptible to prejudice.

References

1. 2013. Learning fair representations. In Proceedings of the 30th International Conference on Machine Learning (ICML13), 325–333.
2. Johannes, R. 1988. Using the adap learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the Annual Symposium on Computer Application in Medical Care, 261. American Medical Informatics Association.
3. Lichman, M. 2013. UCI machine learning repository.
4. Mansouri, K.; Ringsted, T.; Ballabio, D.; Todeschini, R.; and Consonni, V. 2013. Quantitative structure–activity relationship models for ready biodegradability of chemicals. Journal of chemical information and modeling 53(4):867–878.
5. Massart, P. 2007. Concentration inequalities and model selection, volume 6. Springer.
6. Miller, C. C. 2015. Can an algorithm hire better than a human. The New York Times 25.
7. Munoz, C.; Smith, M.; and Patil, D. 2016. Big data: A report on algorithmic systems, opportunity, and civil rights. Executive Office of the President. The White House.
8. Olfat, M., and Aswani, A. 2017. Spectral algorithms for computing fair support vector machines. arXiv preprint arXiv:1710.05895.
9. Paluch, S., and Tuzovic, S. 2017. Leveraging pushed selftracking in the health insurance industry: How do individuals perceive smart wearables offered by insurance organization?
10. Recht, B.; Fazel, M.; and Parrilo, P. A. 2010. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM review 52(3):471–501.
11. Rudin, C. 2013. Predictive policing using machine learning to detect patterns of crime. Wired Magazine, August.
12. Sallis, J.; Bauman, A.; and Pratt, M. 1998. Environmental and policy interventions to promote physical activity a. American journal of preventive medicine 15(4):379–397.
13. Shorack, G. R., and Wellner, J. A. 2009. Empirical processes with applications to statistics, volume 59. SIAM.
14. Smith, J. W.; Everhart, J.; Dickson, W.; Knowler, W.; and
15. Thompson, J. J.; Blair, M. R.; Chen, L.; and Henrey, A. J. 2013. Video game telemetry as a critical tool in the study of complex skill learning. PloS one 8(9):e75129.
16. Tsanas, A., and Xifara, A. 2012. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and Buildings 49:560–567.
17. Wainwright, M. 2017. High-dimensional statistics: A no asymptotic viewpoint.
18. Yeh, I.-C., and Lien, C.-h. 2009. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications 36(2):2473–2480.
19. Zafar, M. B.; Valera, I.; Rodriguez, M. G.; and Gummadi, K. P. 2017. Fairness constraints: Mechanisms for fair classification. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics.
20. Zemel, R.; Wu, Y.; Swersky, K.; Pitassi, T.; and Dwork, C.