NeurIPS Monday

Yuchen Zeng

12/09/2019

1 Joint Affinity Group Poster Session

1.1 Within-and Between-Dimension Clustering Using Regularized Tensor Decomposition

Summary: This work only considers time-location-user tensor. The goal is to perform clustering that is robust to a small shift of time on the observed data tensor. Suppose the original data tensor is X. X^+ is the data tensor whose time moving forward a little bit, compared with X. Similarly, X^- is the tensor whose time moving backward a little bit. They define the penalty as $I = X + w_1 X^+ + w_2 X^-$, where w_1 and w_2 are the weights. They claim that the robustness to time can be achieved by adding the penalty I to the loss function $\min_{rank(\hat{X}) \leq r} ||X - \hat{X}||^2$ (I do not remember the norm). Note the rank here refers to CP rank. They perform clustering on tensor factors on location and users after obtained the low-rank approximation of data tensor.

Comment: The theory is easy to understand. The idea is not bad. It can be applied to many fields if they can extend this idea to other situations instead of being limited in time-location-user data tensor. In my opinion, they can develop more theories in subsequent work, such as computing the boundaries of clustering error.

1.2 Stretching the Effectiveness of MLE from Accuracy to Bias for Pairwise Comparisons

Summary: The goal of this work is to estimate the quality of each item while the observations are comparisons between pairs of items. They consider the Bradley-Terry-Luce (BTL) model. Suppose there are d items of unknown values $\theta^* = [\theta_1^*, ..., \theta_d^*]$, k comparisons per pair of items. The probability that item i beats item j

is $P(i > j) = \frac{e^{\theta_i^*}}{e^{\theta_i^*} + e^{\theta_j^*}} = \frac{1}{1+e^{-(\theta_i^* - \theta_j^*)}}$. Assume θ^* in $\mathcal{D} = \left\{\theta | \sum_{i=1}^{d} \theta_i = 0, ||\theta||_\infty \leq B\right\}$. Then the maximum likelihood estimator (MLE) is defined as $\hat{\theta}_{MLE} = \text{argmax}_{\hat{\theta} \in \mathcal{D}} l(data; \theta)$. They proved that the mean squared error of MLE is $\mathcal{O}(\frac{1}{k})$ and the bias is at least $\Omega(\frac{1}{\sqrt{dk}})$. The definition of their stretched MLE is a little different with that of MLE. The stretched MLE $\hat{\theta}^c_{MLE} = \text{argmax}_{\hat{\theta} \in \mathcal{D}^c} l(data; \theta)$, where $\mathcal{D}^c = \left\{\theta | \sum_{i=1}^{d} \theta_i = 0, ||\theta||_\infty \leq cB\right\}$, c is any constant strictly larger than 1. Their proposed stretched MLE achieves lower bias (I do not remember the specific formula) while maintaining the same mean square error.

Comment: I am confused about the definition of B. They claim that B is a known constant which can be estimated from historical data. In addition, they proposed that the bias is independent with c for all c strictly larger than 1. However, since B is a estimated value, how can we know whether $c\hat{B} > B$?

1.3 Other comments

I am also interested in other two posters that are related to robust classification and metric-space verification, respectively. However, the presenter of robust classification only gave me a brief introduction of its advantages and did not explain to me how it works. The second work shows superior performance compared with previous work in face verification. They did not introduce the theory in depth. I don't really understand how they work.

In addition, I find that generative adversarial networks (GAN) is the most popular topic in today's poster session.

2 Invited talk: Celeste Kidd: How to know

Summary: They investigate learner's certainty in a classic Boolean concept-learning task. They compare different models of certainty in order to determine exactly what learner's subjective certainty judgement encode. They draw five conclusions but I cannot remember clearly (the relationship between surprise and certainty).

Comment: This talk briefly introduced their conclusions and application. I browsed their paper (https://www.celestekidd.com/papers/MartiETAL2016.pdf), but it is very hard to understand.