AIML - CS 337

Lecture 3: ML Terminology

11 August 2022

Lecturer: Abir De *Scribe: Akshay, Alan, Lakshya, Virendra*

Over the last two lectures, we reviewed probability and linear algebra. Now, we introduce some basic ML terminology.

1 An Example

This example is in continuation from Lecture 2. We have a graph $G = (V, E)$, where $V = \{1, \ldots, n\}$ is the set of vertices, and E is the set of edges. The degree of node i is given by d_i. Define the value $x_i(t)$ to be the 'opinion' of node i at time t. $x_1(0), \ldots, x_n(0)$ are given. For $t \geq 0$,

$$x_i(t+1) = \frac{\sum_{j \in \mathcal{N}(i)} x_j(t)}{|\mathcal{N}(i)|} \tag{1}$$

where $\mathcal{N}(i)$ denotes the set of neighbors of node i. That is, the opinion of i at time $t+1$ is the average of the opinions of its neighbors at time t.
Consider $\boldsymbol{x}(t) = [x_1(t) \cdots x_n(t)]^T$. Define $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{A}\boldsymbol{x}(t) = \boldsymbol{x}(t+1)$. From (1), $\boldsymbol{A}$ is a doubly-stochastic matrix, i.e., its entries are non-negative, and its rows and columns add up to 1. For instance, the first row of $\boldsymbol{A}$ is of the form $[a_{1j}]^T$, where $a_{1j} = 1/d_1$ if $j \in \mathcal{N}(i)$, else 0.

We state the following result from [1]. A stochastic matrix $\boldsymbol{M}$ is called semi-positive if all entries of some power $\boldsymbol{M}^\alpha$ are positive.

Theorem 1.1. *If $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ is a semi-positive doubly stochastic matrix, then $\lim_{t \to \infty} \boldsymbol{M}^t = \frac{1}{n}\boldsymbol{J}$, where $\boldsymbol{J} \in \mathbb{R}^{n \times n}$ with all entries 1.*

Thus, if $\boldsymbol{A}$ is semi-positive and doubly stochastic, then for all nodes i, $\lim_{t \to \infty} x_i(t) = \frac{\sum_{j=1}^n x_j(0)}{n}$.

Some points to note:

- $\boldsymbol{A}$ is symmetric
- $\boldsymbol{A}\mathbf{1} = \mathbf{1}$
- $\mathbf{1^T}\boldsymbol{A} = \mathbf{1^T}$

2 Image Classification Problem

We are given a set of images $\{I_1, \ldots, I_n\}$, each corresponding to exactly one of three animal classes: cat, dog, or tiger. The pixel data of I_i is encoded into feature-matrix $X_i \in \mathbb{R}^{d \times d}$, and

IDs of the three classes are $\{0, 1, 2\}$ respectively. *Binary* classification problems involve only two labels, usually $\{0, 1\}$.

The goal is to correctly predict the label of any (unseen) image. We can't possibly write an algorithm for this prediction, since the model itself changes with the input space. Moreover, an algorithm with correct mappings for given set $\{X_i\}$ will perform poorly on unseen images. This is because the model has been completely overfit to these images.

Since we need to train a model using some data, the problem with just the set of X_i's is ill-defined. We also require the labels (y_i's) for a set of examples. This brings the idea of a training set.

3 Training and Validation Sets

3.1 Training Set

For the classification problem, the training set consists of pairs (X_i, y_i). For each image matrix X_i, a label y_i has already been provided, and this label is assumed to be correct. Assuming binary classification, we need to find a function $H : \mathbb{R}^{d\times d} \to \{0, 1\}$, such that $H(X_i) = y_i$. As mentioned above, this function must give correct results for unseen images too.

3.2 Validation Set

How do we know that the algorithm is "ready" for use? Relying only on training set performance isn't a good idea, since the model would have overfit on this data. Thus, we reserve a section of the training data as the validation set. After training the model on the train set, its performance is evaluated on the validation set. Since the latter is unseen during training, we get a more reliable estimate of the model's performance.

References

[1] S. Baik and K. Bang. Limit theorem of the doubly stochastic matrices. *Kangweon-Kyungki Mathematics Journal*, 11(2):155–160, 2003.