Machine Learning: Assignment #2

June 13, 2018

1. Assume a non-uniform PMF for a set of 10 elements. Draw 100 samples from this set with replacement and validate if the distribution of the sample is in alignment with the PMF of the original set.
2. Plot 2D gaussian shapes with different means and covariance matrices. Try the following covariance matrices
   (a) identity matrix * scale
   (b) random matrix
   (c) symmetric matrix
   (d) all ones * scale
3. We saw the method for random sampling and also the cumulative density function for Gaussian distribution. Demonstrate the procedure for random sampling when the data distribution is continuous and Gaussian.
4. We saw that the Gaussian distribution is symmetric. What if your data distribution is non-symmetric and you should model that using Gaussian distributions?
5. There were two parameters for the Gaussian {mean and variance}. How does one affect the other? Does it really affect each other?
6. We saw examples of cardinal variables converted to one-hot encoding. How would you encode ordinals with a lot of values?
7. We saw that the unnormalized variables affect similarity/distance measurements greatly. Are there other ways for handling the shadowing effect besides normalization?
8. We discussed L1 and L2 norms. Find out the meanings for L0 and L\infty norms. What is Frobenius norm?
9. We discussed cosine similarity measure and Euclidean distance measure. How would you choose a metric for measuring the performance of a method or process or a system? When would you use similarity measures and when would you choose distance matrics?

10. Let A and B are data points represented as vectors of size 'p'. What is the interpretation of A.B, A+B, A-B?

11. When vectors are represented in sparse format (libsvm format), define an algorithm for computing the dot product of sparse vectors.

12. We saw that dimension reduction techniques can represent the data in low dimensions without losing much of information. What is the meaning of low dimension representation? If the data can be represented in low dimensions, why are we starting the data representation in higher dimensions? Why not start directly in the low dimension representation itself?

13. Write the expression for matrix multiplication in terms of vector dot products.

14. We saw a technique called basis expansion where we inflated the feature dimensions by combining existing features. This is against the concept of dimension reduction. When would you choose to reduce dimensions vs expand dimensions? Can you do both?

15. We discussed random projections for dimension reduction. Can we do the same for expanding features? Will it help?