Machine Learning Practice and Theory

Day 6 - Unsupervised Learning

Govind Gopakumar

IIT Kanpur

Prelude

Announcements

- Project Groups : Code should be up and running, you should have some idea of what your project is and what the end goals are.
- Quiz 1 is up : Auto graded, feedback. Ask if you have doubts
- Programming assignment 1 is up : Gradient descent

Story so far

Supervised Learning

- KNN, Distance from means
- Decision Trees, Random Forests
- Logistic Regression, Perceptron
- Linear Regression

Techniques

- Gradient Descent
- Formulating a loss function
- Using “maximum probability” to obtain results

Clustering

Clustering - I

Why do we need it?

- Discover patterns or “clusters”
- Preprocessing step for classification
- Allow us to learn “generative” models

What’s the easiest way to do it?

- Group objects together
- But how?

Clustering - II

Model overview

- K-Means clustering : defined by k points
- Each data point is assigned closest mean
- K is sort of a hyper parameter, not to be learnt!

Training the model

- How do we find the means?
- How do we do assignment?
- What are the parameters to be learnt?

Clustering - III

Model parameters

- Known : Location of data
- Unknown : Cluster assignments, cluster means

How to find both?

- Knowing cluster means let us find cluster assignments
- Knowing cluster assignments : does it help the other way around?

Clustering - IV

Alternating optimization

- Two different unknown parameters : μ, z
- Idea from matrix factorization.

Finding the parameters

- How do we do alternating optimization here?
- What are the guesses?
- What does it say about our "loss function"?

Clustering - V

Estimating the cluster IDs

- Iterate over all the means
- Assign cluster ID as the closest mean

Estimating the cluster means

- Collect points belonging to specific cluster
- Compute mean of that cluster

Clustering - VI

Geometry of the model

- Decision surface?
- What sort of clusters does it learn?
- When will it do badly?

Uniqueness of clustering

- What does final cluster depend on?
- Will it always learn good clustering?
- What's an example where it will fail?
- Outliers?

Clustering - VII

Comments about K-Means

- Makes hard assignments
- Size of clusters matters!
- Can work with transformations as well!

Limitations

- Non-convexity of the “loss” function!
- Iterative solution
- Will have to work with better notions of “distance”!
- How do we choose k?

Smarter Clustering

Gaussian Mixture Models - I

Why should we improve our clustering?

- Hard assignment
- Logistic Regression vs other methods!
- Probabilistic interpretation

Generative modelling

- Model how the data was generated!
- Can be used to give new data!
- Preprocessing step for supervised learning?

Gaussian Mixture Models - II

Review of the Gaussian distribution

- p(X) : Reflects how probable a point is
- Density decreases as distance from mean increases
- Variance reflects spread

Estimation of a Gaussian

- Given : A bunch of data points
- What is the most likely Gaussian?
- How do we find it?

Gaussian Mixture Models - III

Modelling assumptions

- Assume each point is “generated” from a Gaussian
- How many Gaussians?
- Where are they?

Model overview

- What are the unknowns in the setting?
- How do we find them?

Gaussian Mixture Models - IV

Alternating optimization?

- Find cluster ID's in a probabilistic sense
- Find clusters also in the same fashion!

What are the parameters then?

- μ for each Gaussian
- Σ for each Gaussian
- Can we make intelligent choices here?

Conclusion

Concluding Remarks

Takeaways

- How to cluster points for unsupervised learning
- How to do alternating optimization (2nd such example)
- Generative modelling and Gaussian Mixture models

Announcements

- Assignment 1, Quiz 1 up.
- (Hopefully programming tutorials also up)

References

- Lecture 10, CS 771 IIT Kanpur
- Lecture 16, CS 771 IIT Kanpur