Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 2

Student Name: Prashant Kumar
Roll Number: 231110036
Date: November 17, 2023

QUESTION 1

The standard the K-means objective function is described as-

$$\mathcal{L}(X, Z, \mu) = \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}||x_n - \mu_k||^2$$

For making k-means online,taking a random example x_n at a time, and then assign x_n "greedily" to the "best" cluster available.

Using the ALT-OPT technique to solve for z_n as follows:
first fix $\mu = \hat{\mu}$ and then solve for z_n

$$\hat{z_n} = arg\ min\sum_{k=1}^{K} z_{nk}||x_n - \hat{\mu}_k||^2 = arg\ min_{z_n} z_{nk}||x_n - \hat{\mu}_{z_n}||^2$$

For performing this step, assigning a cluster to x_n using the mentioned equation, for each of the examples $\{x_n\}_{n=1}^{N}$.

After assigning each example to a cluster, updating the cluster means using the SGD update:

$$\hat{\mu} = arg\ min\mathcal{L}(X, \hat{Z}, \mu) = arg\ min\{\sum_{n=1}^{N}\sum_{n:\hat{Z}_n=k} ||x_n - \mu_k||^2\}$$

$$\hat{\mu_k} = arg\ min_{\mu_k}\{\sum_{n:\tilde{z}_n=k} ||x_n - \mu_k||^2\}$$

At iteration t, we choosing an example x_n uniformly randomly and approximating the gradient g by given formula:

$$g \approx g_n = \frac{\partial}{\partial \mu_k}(||x_n - \mu_k||^2) = -2(x_n - \mu_k)$$

Now, we can update the mean as follows:

$$\mu_k^{(t+1)} = {\mu_k}^{(t)} - \eta g^{(t)}$$

$$\mu_k^{(t+1)} = {\mu_k}^{(t)} + 2\eta({x_n}^{(t)} - {\mu_k}^{(t)})$$

Finally, The step size η can be chosen to be inversely proportional to the number of data points in the kth cluster, i.e., $\eta \propto \frac{1}{N_k}$. This ensures that the updated mean is the centroid of the data points in the cluster. Moreover, there are other ways to define step size such that step size can also be inversely proportional to time which will ensure that centroids are changing faster initially and later the change is slow and convergence is high.

Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 2

Student Name: Prashant Kumar
Roll Number: 231110036
Date: November 17, 2023

QUESTION 2

The objective of maximizing the distance between two classes can be achieved by using the objective loss function given by Fisher's Linear Discriminant Analysis:

$$J(w) = \frac{(w^T\mu_+ - w^T\mu_-)^2}{w^T S_w w} \tag{1}$$

This objective function aims to maximize the separation between the means of the two classes (μ_+ and μ_-), while simultaneously minimizing the separation of data points within each class. It is commonly used in Fisher's Linear Discriminant Analysis for finding an effective one-dimensional projection that enhances class discrimination in binary classification tasks.

where

$$S_w = Sw_1 + Sw_2$$

distance of points in one class and the object function above ensures that inter-class distance will be high and intra-class distance will be less to ensure a large value of $J(w)$

Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 2

Student Name: Prashant Kumar
Roll Number: 231110036
Date: November 17, 2023

QUESTION 3

Assume we have a eigenvector of

$$S = \frac{1}{N} X X^T \tag{2}$$

Now this eigenvector will satisfy the equation $Sv = \lambda v$,where λ is the eigenvalues

$\implies \frac{1}{N}(XX^T)v = \lambda v$

multiplying X^T on both sides- $\implies \frac{1}{N}(X^T X)(X^T v) = \lambda(X^T v)$,

Substitute $u = X^T v$
$\implies \frac{1}{N}(X^T X)(u) = \lambda(u)$

Now notice that, $u = X^T v$ is nothing but an eigenvector for $\frac{1}{N} X X^T$

The advantage of this way of obtaining the eigenvectors is ease of computation.
In normal case, To compute K eigenvector's for $\frac{1}{N} X X^T$ the complexity will be $O(KD^2)$ but in this case the complexity will be $O(KN^2) + O(KND)$

$O(KN^2)$ - for decomposition of $\frac{1}{N} X X^T$
$O(KND)$ - for matrix multiplication
So, The overall complexity will be $O(KND)$ as $N < D$

Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 2

Student Name: Prashant Kumar
Roll Number: 231110036
Date: November 17, 2023

QUESTION 4

(1) A classical linear model is handles linear curve regression. However, this model extends its capability by combining K different linear curves that are weights of K clusters present in the model, it clusters data onto K linear curves and makes predictions for y based on this clustering which results in better prediction as the prediction is dependent on cluster alike data points moreover this approach can effectively reduce outliers in linear curves by potentially isolating them through clustering used in model.

(2) The latent variable model can be defined as follows:

$$p(z_n = k|y_n, \theta) = \frac{p(z_n = k)p(y_n|z_n = k, \theta)}{\sum_{l=1}^{K} p(z_n = l)p(y_n|z_n = l, \theta)}$$
$$p(y_n, z_n|\theta) = p(y_n|z_n, \theta)p(z_n|\theta)$$

Where:

$$p(z_n = k) = \pi_k$$
$$p(y_n|z_n, \theta) = N(w_{z_n}^T x_n, \beta^{-1})$$

ALT-OPT Algorithm

Step 1: Determining the best z_n:

$$z_n = \text{argmax}\, z_n \frac{\pi_k \exp\left(\frac{-\beta}{2}\left(y_n - w{z_n}^T x_n\right)^2\right)}{\sum_{l=1}^{K} \pi_l \exp\left(\frac{-\beta}{2}\left(y_n - w_l^T x_n\right)^2\right)}$$

Step 2: Parameter re-estimation:

$$N_k = \sum_{n=1}^{N} z_{nk}$$
$$w_k = \left(X_k^T X_k\right)^{-1} X_k^T y_k$$
$$\pi_k = \frac{N_k}{N}$$

Here, X_k represents an $N_k \times D$ matrix containing training sets clustered in class k, and y_n denotes $N_k \times 1$ vectors containing training set labels clustered in class k.

When $\pi_k = 1/K$:

$$z_n = \text{argmax}\, z_n \frac{\exp\left(\frac{-\beta}{2}\left(y_n - w{z_n}^T x_n\right)^2\right)}{\sum_{l=1}^{K} \exp\left(\frac{-\beta}{2}\left(y_n - w_l^T x_n\right)^2\right)}$$

This update is similar to multi-output logistic regression that uses K weights.

Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 2

Student Name: Prashant Kumar
Roll Number: 231110036
Date: November 17, 2023

QUESTION 5

Please find the respective images and descriptions in attached python notebooks naemly "231110036-problem-5-part-1.ipynb", "231110036-problem-5-part-2.ipynb" and "231110036-problem-5-part-3.ipynb"

Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 2

Student Name: Prashant Kumar
Roll Number: 231110036
Date: November 17, 2023

QUESTION 6

My solution to problem 6