Machine Learning (CS 6140) Homework 3

Xiao Wang

November 26, 2017

Problem 1

$$L(u, \mu, \{x_i\}, \gamma, \Lambda) = \frac{1}{2}\sum_{i=1}^{N} \|y_i - (ux_i + \mu)\|_2^2 + \gamma^\top \sum_{i=1}^{N} x_i + tr[\Lambda^\top (u^\top u - I_d)]$$
$$\frac{\partial L}{\partial u} = 0 \implies \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i$$
$$\frac{\partial L}{\partial x_i} = 0 \implies x_i = u^\top (y_i - \hat{\mu})$$
$$\implies \max tr(YY^\top uu^\top), \text{ s.t. } uu^\top = I_d$$
$$YY^\top = u_y \Sigma_y^2 u_y^\top, uu^\top = u_B \Sigma_B^2 u_B^\top$$

Using Von Neumann's Inequality, $u_y = u_B$

$$uu^\top = \begin{bmatrix} u & \text{noise} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \text{noise} \end{bmatrix} \begin{bmatrix} u^\top \\ \text{noise}^\top \end{bmatrix}$$
$$\Sigma_B = diag(\sigma_1, \cdots, \sigma_N)$$

Because $\sigma_1, ..., \sigma_N$ are in descending order. Thus, we only need $\sigma_{1..d}$ to preserve the significant eigenvectors. Also $\lambda_i = \sigma_i^2$.Therefore, the objective function to minimize can be express as $\sum_{i \geq d+1} \sigma^2$.

Problem 2

a)

b) **Step 1:**
- Compute distance between each pair of data points: $O(DN^2)$
- Select k nearest neighbors for each point with quick select algorithm: $O(N^2)$

Overall time complexity: $O(DN^2)$

Step 2:
- Multiply $k \times D$ matrix and $D \times k$ matrix: $O(Dk^2)$
- Each point need one multiplication Overall time complexity: $O(DNk^2)$

Problem 3

a)

$$
\begin{aligned}
x^\top L x &= \begin{bmatrix} x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} D_{11} & -W_{12} & \cdots & -W_{1N} \\ -W_{21} & D_{22} & \cdots & -W_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -W_{N1} & -W_{N2} & \cdots & D_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \\
&= D_{11}x_1^2 - x_1 \sum_{i=1}^{N} W_{i1}x_i + \cdots + D_{NN}x_N^2 - x_N \sum_{i=1}^{N} W_{iN}x_i \\
&= x_1 \sum_{i=1}^{N} W_{1i}(x_1 - x_i) + \cdots + x_N \sum_{i=1}^{N} W_{Ni}(x_N - x_i) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}(x_i - x_j)^2
\end{aligned}
$$

b) From Question a), we can see that for any x, we have $x^\top Lx \geq 0$. Thus L is positive semi-definite.

c) L is not invertible. Because $\mathbf{1}$ is an eigenvector of L with eigenvalue 0.

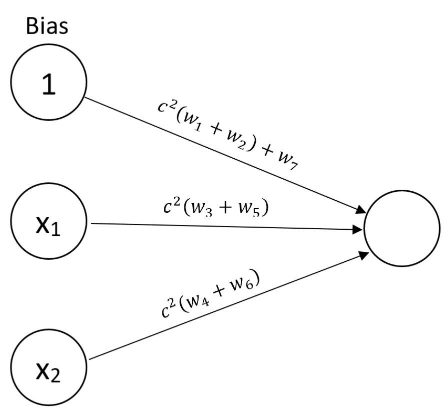
d) N(zero singular value)=N(connected components in graph)
e.g. Suppose vertex 1...m is a connected component, then [1 ...(k)... 1 0 ...] is an eigenvector of L with eigenvalue 0.

Problem 4

a)

$$P(y = 1|x, w) = [1 + e^{-[c^2(w_3+w_4)x_1+c^2(w_5+w_6)x_2+c^2(w_1+w_2)+w_7]}]^{-1}$$

If $[c^2(w_3 + w_4)x_1 + c^2(w_5 + w_6)x_2 + c^2(w_1 + w_2) + w_7] > 0$, the prediction is $y = 1$, otherwise $y = 0$.



b)

c) **True.**
Because the linear function of a linear function is still a linear function of the original input.