ANLY561 Homework 9

Hongyang Zheng

Question 1

Let $f(\mathbf{x}) = ||\mathbf{y} - \mathbf{x}||_2^2$ and $g(\mathbf{x}) = \mathbf{v^T x} - b$

Then we have $\nabla f(\mathbf{x}) = 2(\mathbf{x} - \mathbf{y})$ and $\nabla g(\mathbf{x}) = \mathbf{v}$

Since $\mathbf{v}$ is nonzero all the time, there is always a λ such that $\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x})$

Then we have $2(\mathbf{x} - \mathbf{y}) = \lambda \mathbf{v}$ and $\mathbf{x} = \frac{\lambda \mathbf{v}}{2} + \mathbf{y}$

Plug $\mathbf{x}$ into $\mathbf{v^T x} - b = 0$, we get $\mathbf{v^T}(\frac{\lambda \mathbf{v}}{2} + \mathbf{y}) - b = 0$, therefore $\lambda = \frac{2}{||\mathbf{v}||_2^2}(b - \mathbf{v^T y})$

$$\mathbf{x} = \frac{\lambda \mathbf{v}}{2} + \mathbf{y} = \frac{\frac{2\mathbf{v}}{||\mathbf{v}||_2^2}(b - \mathbf{v^T y})}{2} + \mathbf{y} = \frac{\mathbf{v}}{||\mathbf{v}||_2^2}(b - \mathbf{v^T y}) + \mathbf{y}$$

$$\mathbf{x}^* = \frac{\mathbf{v}}{||\mathbf{v}||_2^2}(b - \mathbf{v^T y}) + \mathbf{y}$$

Question 2

a)

We know that $\mathbf{x}_i$ is the ith row of data matrix X and it has d columns.

Therefore, $\mathbf{x_i} = \begin{pmatrix} X_{i,1} \\ X_{i,2} \\ . \\ . \\ . \\ X_{i,d} \end{pmatrix}$

$$
\begin{aligned}
\sum_{i=1}^{n} ||\mathbf{x}_i - c_i\mathbf{u}||_2^2 &= \sum_{i=1}^{n} (\sqrt{|\mathbf{x}_i - c_i\mathbf{u}|})^2 \\
&= \sum_{i=1}^{n} |\mathbf{x}_i - c_i\mathbf{u}| \\
&= \sum_{i=1}^{n}((X_{i,1} - c_i\mathbf{u_1})^2 + (X_{i,2} - c_i\mathbf{u_2})^2 + \ldots + (X_{i,d} - c_i\mathbf{u_d})^2) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{d}(X_{i,j} - c_i\mathbf{u_j})^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{d}(X - \mathbf{cu^T})_{i,j}^2 \\
&= ||X - \mathbf{cu^T}||_{Fro}^2
\end{aligned}
$$

b)

Let $f(c) = ||\mathbf{x_i} - c_i\mathbf{u}||_2^2$.

Then we have $\nabla f(c) = 2\mathbf{u^T}(c_i\mathbf{u} - \mathbf{x_i})$ and $\nabla^2 f(c) = ||\mathbf{u}||_2^2$, therefore, $f(c)$ is strictly convex, which implies the program has a unique solution.

Let $\nabla f(c) = 0$, then we have $c^* = \mathbf{u^T}\mathbf{x_i} = \mathbf{x_i^T}\mathbf{u}$

Therefore, $c^* = \mathbf{x_i^T}\mathbf{u}$ is the unique solution to minimize $f(c)$.

c)

From part a) we know that $\sum_{i=1}^{n} ||\mathbf{x}_i - c_i\mathbf{u}||_2^2 = ||X - \mathbf{cu^T}||_{Fro}^2$

From part b) we know that $c^* = \mathbf{x_i^T}\mathbf{u}$ is the unique solution to minimize $||\mathbf{x_i} - c_i\mathbf{u}||_2^2$

In other words, $c_i^* = \mathbf{x_i^T}\mathbf{u}$ is the solution to minimize $||\mathbf{x_i} - c_i\mathbf{u}||_2^2$, where $\mathbf{x_i}$ is the ith row of matrix X.

Let $\mathbf{c}^* = \begin{pmatrix} c_1 \\ c_2 \\ . \\ . \\ . \\ c_n \end{pmatrix}$

Therefore, $\mathbf{c}^* = X\mathbf{u}$ is the solution to minimize $\sum_{i=1}^{n} ||\mathbf{x}_i - c_i\mathbf{u}||_2^2$

Therefore, $\mathbf{c}^* = X\mathbf{u}$ is the solution to minimize $||X - \mathbf{cu^T}||_{Fro}^2$

Plug $\mathbf{c}^*$, we get $\sum_{i=1}^{n} ||\mathbf{x}_i - c_i\mathbf{u}||_2^2 = ||X - X\mathbf{uu^T}||_{Fro}^2$

With the same constraint, the two program are same.

d)

To simplify the process, any value that is not on the diagonal entries of A^TA is represented by a.

Matrix $A \in M_{m,n}$ and $A^TA \in M_{n,n}$

$$A^TA = \begin{pmatrix} \sum_{i=1}^{m} A_{i,1}^2 & . & . \; . & a \\ a & \sum_{i=1}^{m} A_{i,2}^2 & . \; . & a \\ . & . & & . \\ . & & . & . \\ . & & & . \\ a & . & . \; . & \sum_{i=1}^{m} A_{i,n}^2 \end{pmatrix}$$

Therefore,
$trace(A^TA) = \sum_{i=1}^{m} A_{i,1}^2 + \sum_{i=1}^{m} A_{i,2}^2 + \ldots + \sum_{i=1}^{m} A_{i,n}^2 = \sum_{i=1}^{m}(A_{i,1}^2 + A_{i,2}^2 + \ldots + A_{i,n}^2) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j}^2$

Since $||A||_{Fro}^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j}^2$, $||A||_{Fro}^2 = trace(A^TA)$

$$A^TB = \begin{pmatrix} \sum_{i=1}^{m} A_{i,1}B_{i,1} & . & . \; . & a \\ a & \sum_{i=1}^{m} A_{i,2}B_{i,2} & . \; . & a \\ . & . & & . \\ . & & . & . \\ . & & & . \\ a & . & . \; . & \sum_{i=1}^{m} A_{i,n}B_{i,n} \end{pmatrix}$$

$trace(A^TB) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j}B_{i,j}$

$$BA^T = \begin{pmatrix} \sum_{j=1}^{n} A_{1,j}B_{1,j} & . & . \; . & a \\ a & \sum_{j=1}^{n} A_{2,j}B_{2,j} & . \; . & a \\ . & . & & . \\ . & & . & . \\ . & & & . \\ a & . & . \; . & \sum_{j=1}^{n} A_{m,j}B_{m,j} \end{pmatrix}$$

$trace(BA^T) = \sum_{j=1}^{n}(A_{1,j}B_{1,j} + A_{2,j}B_{2,j} + \ldots + A_{m,j}B_{m,j}) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j}B_{i,j}$

Therefore, $trace(A^TB) = trace(BA^T)$

e)

$$
\begin{aligned}
\min_{u \in R^d} ||X - X\mathbf{u}\mathbf{u}^{\mathbf{T}}||_{Fro}^2 &= \min_{u \in R^d} trace((X - X\mathbf{u}\mathbf{u}^{\mathbf{T}})(X - X\mathbf{u}\mathbf{u}^{\mathbf{T}})^T) \\
&= \min_{u \in R^d} trace((X - X\mathbf{u}\mathbf{u}^{\mathbf{T}})(X^T - \mathbf{u}\mathbf{u}^{\mathbf{T}}X^T)) \\
&= \min_{u \in R^d} trace(XX^T - X\mathbf{u}\mathbf{u}^{\mathbf{T}}X^T - X\mathbf{u}\mathbf{u}^{\mathbf{T}}X^T + X\mathbf{u}\mathbf{u}^{\mathbf{T}}\mathbf{u}\mathbf{u}^{\mathbf{T}}X^T) \\
&= \min_{u \in R^d} trace(XX^T - X\mathbf{u}\mathbf{u}^{\mathbf{T}}X^T) \\
&= \min_{u \in R^d} trace(XX^T - \mathbf{u}^TX^TX\mathbf{u}) \\
&= \min_{u \in R^d} (trace(XX^T) - trace(\mathbf{u}^TX^TX\mathbf{u})) \\
&= trace(XX^T) - \max_{u \in R^d} trace(\mathbf{u}^TX^TX\mathbf{u}) \\
&= \max_{u \in R^d} (\mathbf{u}^TX^TX\mathbf{u})
\end{aligned}
$$

Therefore, the minimization program and the maximization program are equivalent.

f)

The program

$$\min_{u \in R^d, c \in R^n} \sum_{i=1}^{n} ||\mathbf{x}_i - c_i\mathbf{u}||_2^2$$

subject to $||\mathbf{u}||_2^2 = 1$

is equivalent to the program

$$\min_{u \in R^d, c \in R^n} ||X - X\mathbf{u}\mathbf{u}^{\mathbf{T}}||_{Fro}^2$$

subject to $||\mathbf{u}||_2^2 = 1$

and then is equivalent to the program

$$\max_{u \in R^d} (\mathbf{u}^TX^TX\mathbf{u})$$

subject to $||\mathbf{u}||_2^2 = 1$

From part c), we know that $\mathbf{c}^* = X\mathbf{u}$ can solve the program

$$\min_{c \in R^n} \sum_{i=1}^{n} ||\mathbf{x}_i - c_i\mathbf{u}||_2^2$$

subject to $||\mathbf{u}||_2^2 = 1$, therefore, $\mathbf{c}^*$ is part of the solution for the program

$$\max_{u \in R^d} (\mathbf{u}^TX^TX\mathbf{u})$$

subject to $||\mathbf{u}||_2^2 = 1$

Since $\mathbf{u}^*$ is a normalized eigenvector of X^TX corresponding the largest eigenvalue of X^TX, then $\mathbf{u}^*$ is a solution for the program $\max_{u \in R^d} (\mathbf{u}^TX^TX\mathbf{u})$subject to $||\mathbf{u}||_2^2 = 1$

Plug $\mathbf{u}^*$ into $\mathbf{c}^* = X\mathbf{u}$, we have $\mathbf{c}^* = X\mathbf{u}^*$. Therefore, $\mathbf{c}^*$ and $\mathbf{u}^*$ together solve the program $\max_{u \in R^d} (\mathbf{u}^TX^TX\mathbf{u})$subject to $||\mathbf{u}||_2^2 = 1$, which is equivalent to the program $\min_{u \in R^d, c \in R^n} \sum_{i=1}^{n} ||\mathbf{x}_i - c_i\mathbf{u}||_2^2$