DDA3020 Assignment 1 -- Written & Report

Xue Zhongkai 122090636

I. Written Problems

Problem 1

1. $R^T R = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$= \begin{pmatrix} \cos^2\theta + \sin^2\theta & -\sin\theta\cos\theta + \cos\theta\sin\theta & 0 \\ -\cos\theta\sin\theta + \sin\theta\cos\theta & (-\sin\theta)^2 + \cos^2\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$= I$

=> Thus R is an orthogonal matrix.

2. As an orthogonal matrix,

$Q^TQ = QQ^T = I$

Let (λ, v) be the eigen-pair s.t.

$Qv = \lambda v \quad (v \neq 0)$

We get $Q^TQv = Q^T\lambda v$

$v = \lambda Q^T v$

Take the dot product with v on both sides,

$\|v\|^2 = \lambda (Q^T v)^T \cdot v = \lambda v^T \cdot (Qv) = \lambda v^T(\lambda v) = \lambda^2 \|v\|^2$

$\|v\| \neq 0$ since v is not a zero vector, thus $\lambda^2 = 1$

Under the circumstance where $\lambda \in \mathbb{R}$, $\lambda = \pm 1$.

Problem 2

(1) By def. $\forall x_1, x_2, \lambda \in (0,1)$

$|\lambda x_1 + (1-\lambda)x_2| \le \lambda|x_1| + (1-\lambda)|x_2|$.

which directly comes out from Triangle Inequality.

(2) We have the gradient to be $\nabla f(x) = 2A^T(Ax-b)$

and the Hessian $\nabla^2 f(x) = 2A^TA$

$A^TA \succeq 0$ hence $\nabla^2 f(x) \succeq 0$

By def $f(x) = \|Ax-b\|^2$ is convex.

Problem 3

Recall that $H_P(X) = -\sum_x P(x)\log P(x)$

$H_{P,Q}(X) = -\sum_x P(x)\log Q(x)$

Apply Jensen's inequality on $\log(\cdot)$,

$\log\left(E\left[\frac{Q(x)}{P(x)}\right]\right) \ge E\left[\log\left(\frac{Q(x)}{P(x)}\right)\right]$

LHS: $\log\left(\sum_x P(x)\frac{Q(x)}{P(x)}\right) = \log\sum_x Q(x) = \log 1 = 0$

RHS: $\sum_x P(x)\log\left(\frac{Q(x)}{P(x)}\right)$

Hence, $0 \ge \sum_x P(x)\log\left(\frac{Q(x)}{P(x)}\right) = \sum_x P(x)(\log Q(x) - \log P(x))$

$-\sum_x P(x)\log P(x) \le -\sum_x P(x)\log Q(x)$

i.e. $H_P(x) \le H_{P,Q}(x)$, "=" only when $P(x) = Q(x)$

Problem 4

(1) To simplify, combine the intercept b into the weights w_i,

then the extended $\tilde{x}_i = [1, x_i^T]^T$, $w = [b, w_i]^T$

The design matrix appears as $\tilde{X} = \begin{bmatrix} \text{---}\tilde{x}_1\text{---} \\ \text{---}\tilde{x}_2\text{---} \\ \vdots \\ \text{---}\tilde{x}_n\text{---} \end{bmatrix}$

The problem converts to

$\min_w \sum_{i=0}^N (\tilde{x}_i^T w - y_i)^2 + \lambda \tilde{w}^T\tilde{w}$

In the form of matrices, it is

$\min_w (\tilde{X}W - y)^T(\tilde{X}W - y) + \lambda\tilde{w}^T\tilde{w}$

Set $\frac{\partial}{\partial w}(\tilde{X}W-y)^T(\tilde{X}W-y) + \lambda\tilde{w}^T\tilde{w} = 0$

$2\tilde{X}^T\tilde{X}W - 2\tilde{X}^Ty + 2\lambda\hat{I}_d W = 0$

$\Rightarrow W = (\tilde{X}^T\tilde{X} + \lambda\hat{I}_d)^{-1}\tilde{X}^Ty$

(2) Let $J(w,b) = \sum_{i=1}^N (f_{w,b}(x_i) - y_i)^2 + \lambda\tilde{w}^T\tilde{w}$

$= \sum_{i=1}^N (x_i^Tw + b - y_i)^2 + \lambda\tilde{I}w^Tw$

where $\tilde{I} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$

Consider one-example case first for simplicity.

$\frac{\partial}{\partial w}J(w,b) = 2(x_i^Tw + b - y_i)\cdot x_i + 2\lambda\tilde{I}$

$\frac{\partial}{\partial b}J(w,b) = 2(x_i^Tw + b - y_i)$

Rewrite it as the summation in the matrix,

$w := w + \alpha\left(2\sum_{i=1}^N (x_i^Tw + b - y_i)x_i + 2\lambda\tilde{I}\right)$

$b := b + \alpha\left(2\sum_{i=1}^N (x_i^Tw + b - y_i)\right)$

Initialize a specific w_0 and b_0, choose an appropriate learning rate α and iterate until converging to some extent.

Problem 5

Given the training data, we have the design and response as

$\Phi = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$, $Y = \begin{bmatrix} -1 & -1 \\ -1 & -2 \\ -2 & -1 \\ 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}$

Set $\varepsilon_i = y_i - \hat{y}_i \sim \mathcal{N}(0, \sigma^2)$,

$l(\theta) = \log L(\theta)$

$= \log\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_i - w_i^T\phi_i(x))^2}{2\sigma^2}\right)$

$= \sum_{i=1}^n \log\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_i - w_i^T\phi_i(x))^2}{2\sigma^2}\right)$

$= n\log\frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2}\cdot\frac{1}{2}\sum_{i=1}^n (y_i - w_i^T\phi_i(x))^2$

Hence maximizing this log-likelihood is equivalent to

$\min_w \sum_{i=1}^N (\hat{y}_i - y_i)^2 = \sum_{i=1}^N (W^T\phi(x_i) - y_i)^2$

$\frac{\partial}{\partial w}(W^T\phi(x_i) - y_i)^2 = 2(W^T\phi(x_i) - y_i)^T\phi(x_i) = 0$

$\Rightarrow W = (\Phi^T\Phi)^{-1}\Phi^TY = \begin{bmatrix} -\frac{4}{3} & -\frac{4}{3} \\ \frac{4}{3} & \frac{4}{3} \end{bmatrix}$