HW1 (Math)

1. $X = \left[\begin{array}{c} \underbrace{\boxed{\quad\quad\quad}}_{} \\ \\ \\ \end{array} \right] m = \sum_{i=1}^{r} s_i u_i v_i^T$

$\overbrace{\quad\quad\quad}^{d}$

(a) $X = \begin{bmatrix} e_1 \\ \vdots \\ e_1 \\ e_2 \\ \vdots \\ e_2 \\ \vdots \\ e_n \\ \vdots \\ e_n \end{bmatrix} \begin{array}{l} \} n_1 \\ \\ \\ \} n_2 \\ \\ \\ \\ \} n_n \\ \\ \end{array} \leftarrow y_{ij} \Big]_{j=1}^{n_i}$

Empirical Risk: $\frac{1}{2n} \| Xw - y \|^2$

$$\nabla \hat{R}(w) = \frac{1}{2n} X^T (Xw - y) = 0$$

so $Xw - y = 0$

$$Xw = y$$

$$\begin{bmatrix} e_1 \\ \vdots \\ e_1 \\ e_2 \\ \vdots \\ e_2 \\ \vdots \\ e_n \\ \vdots \\ e_n \end{bmatrix} \begin{array}{l} \} n_1 \\ \\ \\ \} n_2 \\ \\ \\ \\ \} n_d \\ \\ \end{array} \begin{bmatrix} w_1 \\ \vdots \\ \vdots \\ w_d \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \end{bmatrix} \\ \\ \begin{bmatrix} y_{n1} \\ y_{nn_d} \end{bmatrix} \end{bmatrix}$$

$$n_j w_j = \sum_{j=1}^{n_i} y_{ij}$$

$$w_j = \frac{1}{n_j} \sum_{j=1}^{n_i} y_{ij}$$

(b) Suppose $y = \sum_{i=1}^{r} a_i u_i \qquad a_i \in R$

$Xw = y$ is the optimal case.

So $w = w_{ols} \qquad Xw = Xw_{ols} = XX^+ y$

$$X = \sum_{i=1}^{r} s_i u_i v_i^T$$

$$X^+ = \sum_{i=1}^{r} \frac{1}{s_i} v_i u_i^T$$

$$XX^+ y = \sum_{i}^{r} \sum_{j}^{r} \sum_{k}^{r} s_i u_i \underbrace{v_i^T v_j u_j^T u_k}_{\substack{\text{when } i=j=k \\ \text{not satisfied} \\ =0}} a_k$$

$$\Rightarrow XX^+ y = \sum_{i}^{r} u_i a_i = y$$

So $Xw = y$

(c) Prove X^TX full rank $\Leftrightarrow$ X full rank

$\equiv$ $Null(X^TX)=\{0\} \Leftrightarrow Null(X)=\{0\}$

($\Rightarrow$) Proof by contradiction (MML 4.71)

X not full rank $\Rightarrow$ X^TX not full rank.

Suppose X not invertible,

then singular value has some zeros.

Then singular value of X^TX has 0 included. So X^TX not full rank

---

Suppose X invertible $\Rightarrow$ $N(X)=\{0\}$

($\Leftarrow$) Let $\vec{v} \in N(X^TX)$

So $X^TX\vec{v}=0$

So $\|Xv\|_2^2 = v^T \underbrace{X^TXv}=0$

So $X\vec{v}=0$ so $\vec{v}=\{\vec{0}\}$

So X^TX invertible.

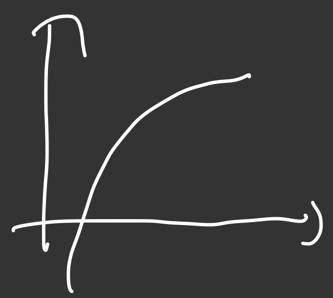
(d) $X = \begin{bmatrix} 1 \\ 1 \end{bmatrix}^T$

$X^T X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ det = 0 singular.

$X X^T = [2]$ det $= 2 \neq 0$

invertible

2. (a) log



so $x > y$ implies $\log(x) > \log(y)$

when $x > y > 0$.

probability are all non-negative.

so optimize log-likelihood is same as optimize original likelihood.

so $\arg\max_w L(w) = \arg\max \log L(w)$

$$= \underset{w}{\arg\max} \sum_{i=1}^{n} \sum_{j=1}^{k} y_{ij} \log(P_w(t_j | x_i))$$

Also minimize the negative version same as maximize original one.

$$\text{So } RHS = \underset{w}{\arg\min} - \sum_{i=1}^{n} \sum_{j=1}^{k} y_{ij} \log(P_w(t_j | x_i))$$

(b) $-\log(P_w(1|x)) = \log(\frac{1}{P_w(1|x)}) = \ln \frac{1}{P_w(1|x)}$

$-\log(1 - P_w(1|x)) = \log(\frac{1}{1-P_w(1|x)}) = \ln \frac{1}{P_w(-1|x)}$

(c) $\log(\prod_{i=1}^{n} P_w(y_i | x_i) P(w))$

$$= \sum_{i=1}^{n} \log(P_w(y_i|x_i)\ P(w))$$

$$= \sum_{i=1}^{n} \log P_w(y_i | x_i) + \log P(w)$$

$$= \sum_{i=1}^{n} \cancel{\frac{1}{\sqrt{2\pi}}} (-\frac{1}{2}(y_i - w^T x_i)^2 + \sum_{j=1}^{d} \cancel{\frac{\lambda}{\sqrt{2\pi}}} (-\frac{1}{2}) w_j^2$$

$$= -\frac{1}{2} \sum_{i=1}^{n} (y_i - w^T x_i)^2 + \sum_{j=1}^{d} (-\frac{\lambda}{2})(w_j^2)$$

negative this, we set.

$$\underset{w}{\arg\min}\left\{\frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i)^2 + \frac{\eta}{2}\sum_{j=1}^{d} w_j^2\right\}$$

4. (a) $\phi(x) = (1, x_1, \ldots\ldots x_d, x_1^2 \ldots\ldots x_d^2$

$x_1x_2 \ldots\ldots x_1x_d; \ldots, x_{d-1}x_d)$

$\phi(x) = (1, x_1, x_2, x_3, x_1^2, x_2^2, x_3^2$

$x_1x_2, x_1x_3, x_2x_3)$

5. $$\hat{R}_{\log}(w) = \frac{1}{n}\sum_{i=1}^{n} \ln(1+\exp(-y_i w^T x_i))$$

$$\nabla_w \hat{R}_{\log}(w) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+\exp(-y_i w^T x_i)} \cdot (\exp(-y_i w^T x_i))'$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+\exp(-y_i w^T x_i)} \cdot (\exp(-y_i w^T x_i) \cdot (-y_i x_i))$$