Machine Learning

Answer Sheet for Homework 0

Da-Min HUANG

B00502124

Department of Electrical Engineering, National Taiwan University, Taipei 106, Taiwan

December 8, 2015

1 Probability and Statics

Problem 1

Now we have

1. $C(N,\ 0) = C(N,\ N) = 1$
2. $C(N,\ K) = C(N-1,\ K) + C(N-1,\ K-1)$

for $N \geq 1$, $N \geq K \geq 0$.
Claim:

$$C(N,\ K) = \frac{N!}{K!\,(N-K)!} \tag{1}$$

Proof of Claim:
Prove by induction:

$$C(N-1,\ K) + C(N-1,\ K-1) = \frac{(N-1)!}{K!\,((N-1)-K)!} + \frac{(N-1)!}{(K-1)!\,((N-1)-(K-1))!} \tag{2}$$

$$= \frac{(N-K)(N-1)! + K(N-1)!}{K!\,(N-K)!} \tag{3}$$

$$= \frac{N(N-1)!}{K!\,(N-K)!} = \frac{N!}{K!\,(N-K)!} = C(N,\ K) \tag{4}$$

$\square$

Problem 2

The probability of getting exactly 4 heads is

$$P_{4\text{ heads}} = \binom{10}{4}(0.5)^{10} = \frac{105}{512} \tag{5}$$

The probability of getting full house is

$$P_{\text{full house}} = \binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}\bigg/\binom{52}{5} = \frac{6}{4165} \tag{6}$$

$\square$

Problem 3

$$P = \frac{1}{\binom{3}{1}+\binom{3}{2}+\binom{3}{3}} = \frac{1}{7} \tag{7}$$

$\square$

Problem 4

$$P = \frac{\frac{1}{2}\times\frac{1}{4}}{\frac{1}{2}\times\frac{1}{8}+\frac{1}{2}\times\frac{1}{4}} = \frac{2}{3} \tag{8}$$

$\square$

Problem 5

$$\max\left(P\left(A\cap B\right)\right) = 0.3 \tag{9}$$

$$\min\left(P\left(A\cap B\right)\right) = 0 \tag{10}$$

$$\max\left(P\left(A\cup B\right)\right) = 0.7 \tag{11}$$

$$\min\left(P\left(A\cup B\right)\right) = 0.4 \tag{12}$$

$$\tag{13}$$

$\square$

Problem 6

We have

$$\left(X_n - \overline{X}\right)^2 = X_n^2 - 2X_n\overline{X} + \overline{X}^2 \tag{14}$$

so

$$\begin{aligned}
\sigma_X^2 &= \frac{1}{N-1}\sum_{n=1}^{N}\left(X_n^2 - 2X_n\overline{X} + \overline{X}^2\right) && (15)\\
&= \frac{1}{N-1}\left(\sum_{n=1}^{N} X_n^2\right) - \frac{2\overline{X}}{N-1}\left(\sum_{n=1}^{N} X_n\right) + \frac{N\overline{X}^2}{N-1} && (16)\\
&= \frac{1}{N-1}\left(\sum_{n=1}^{N} X_n^2\right) - \frac{2N\overline{X}^2}{N-1} + \frac{N\overline{X}^2}{N-1} && (17)\\
&= \frac{N}{N-1}\left[\left(\frac{1}{N}\sum_{n=1}^{N} X_n^2\right) - \overline{X}^2\right] && (18)
\end{aligned}$$

□

Problem 7

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_{X_2}(z - x_1)\, f_{X_1}(x_1)\, dx_1 && (19)\\
&= \int_{-\infty}^{+\infty}\left[\frac{1}{\sqrt{2\pi}\sigma_{X_2}}\exp\left(-\frac{(z - x_1 - \mu_{X_2})^2}{2\sigma_{X_2}^2}\right)\right]\left[\frac{1}{\sqrt{2\pi}\sigma_{X_1}}\exp\left(-\frac{(x_1 - \mu_{X_1})^2}{2\sigma_{X_1}^2}\right)\right] dx_1 && (20)\\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2}}\exp\left[-\frac{\left(z - (\mu_{X_1} + \mu_{X_2})^2\right)}{2\left(\sigma_{X_1}^2 + \sigma_{X_2}^2\right)}\right] && (21)
\end{aligned}$$

So the $\mu_Z = -1$ and $\sigma_Z^2 = 5$.

□

---

2 Linear Algebra

Problem 1

The r.r.e.f. of this matrix is

$$\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \tag{22}$$

so the rank is 2.

$\square$

Problem 2

$$\frac{1}{8}\begin{pmatrix} 1 & -5 & 6 \\ -2 & 6 & -4 \\ 3 & -3 & 2 \end{pmatrix} \tag{23}$$

$\square$

Problem 3

Consider

$$\det\begin{pmatrix} 3-\lambda & 1 & 1 \\ 2 & 4-\lambda & 2 \\ -1 & -1 & 1-\lambda \end{pmatrix} = -\lambda^3 + 8\lambda^2 - 20\lambda + 16 \tag{24}$$

$$= -\left(x-4\right)\left(x-2\right)^2 \tag{25}$$

So the eigenvalues are

$$\lambda_1 = 4,\ \ \lambda_2 = \lambda_3 = 2 \tag{26}$$

Then,

$$\begin{cases} -x_1 + x_2 + x_3 = 0 \\ 2x_1 + 2x_3 = 0 \\ -x_1 - x_2 - 3x_3 = 0 \end{cases},\ \begin{cases} x_1 + x_2 + x_3 = 0 \\ 2x_1 + 2x_2 + 2x_3 = 0 \\ -x_1 - x_2 - x_3 = 0 \end{cases} \tag{27}$$

The eigenvectors are

$$\mathbf{v}_1 = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix},\ \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix},\ \mathbf{v}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \tag{28}$$

$\square$

Problem 4

It is easy to show that $\Sigma\Sigma^\dagger = I$ or $\Sigma^\dagger\Sigma = I$. So if Σ is invertible, then $\Sigma^\dagger = \Sigma^{-1}$.

Now we have $M = U\Sigma V^T$, so

$$\begin{aligned} MM^\dagger M = \left(U\Sigma V^T\right)\left(V\Sigma^\dagger U^T\right)\left(U\Sigma V^T\right) &= U\left(\Sigma V^T V\Sigma^\dagger U^T U\Sigma\right)V^T && (29) \\ &= U\left[\Sigma\left(V^T V\right)\Sigma^\dagger\left(U^T U\right)\Sigma\right]V^T = U\left(\Sigma\Sigma^\dagger\Sigma\right)V^T && (30) \\ &= U\left(I\Sigma\right)V^T = U\Sigma V^T && (31) \end{aligned}$$

where

$\square$

Problem 5

Let $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$, $Z = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$, so

$$\begin{aligned} \mathbf{x}^T Z Z^T \mathbf{x} &= \begin{pmatrix} \sum_{i=1}^m x_i a_{i1} & \sum_{i=1}^m x_i a_{i2} & \cdots & \sum_{i=1}^m x_i a_{in} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^m x_i a_{i1} \\ \sum_{i=1}^m x_i a_{i2} \\ \vdots \\ \sum_{i=1}^m x_i a_{in} \end{pmatrix} && (32) \\ &= \sum_{j=1}^n \left(\sum_{i=1}^m x_i a_{ij}\right)^2 \geq 0 && (33) \end{aligned}$$

If now A is PD, then

$$\begin{aligned} &\mathbf{x}^T A\mathbf{x} > 0 && (34) \\ &\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T A^T \mathbf{x} = (A\mathbf{x})^T \mathbf{x} = (\lambda\mathbf{x})^T \mathbf{x} = \lambda\|\mathbf{x}\|^2 > 0 && (35) \end{aligned}$$

so $\lambda > 0$. If all eigenvalues are all positive, then

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T A^T \mathbf{x} = (A\mathbf{x})^T \mathbf{x} = (\lambda\mathbf{x})^T \mathbf{x} = \lambda\|\mathbf{x}\|^2 > 0 \quad (36)$$

so A is PD.

$\square$

Problem 6

$$\max\left(\mathbf{u}^T\mathbf{x}\right) = \|\mathbf{x}\| \Rightarrow \mathbf{u} = \frac{\mathbf{x}}{\|\mathbf{x}\|} \quad (37)$$

$$\min\left(\mathbf{u}^T\mathbf{x}\right) = -\|\mathbf{x}\| \Rightarrow \mathbf{u} = -\frac{\mathbf{x}}{\|\mathbf{x}\|} \tag{38}$$

$$\min\left(\left|\mathbf{u}^T\mathbf{x}\right|\right) = 0 \Rightarrow \mathbf{u} \perp \mathbf{x} \tag{39}$$

$\square$

Problem 7

$$\left|\mathbf{w}^T\left(\mathbf{x}_1 - \mathbf{x}_2\right)\right| = \left|\mathbf{w}^T\mathbf{x}_1 - \mathbf{w}^T\mathbf{x}_2\right| = 5 \tag{40}$$

$\square$

---

3 Calculus

Problem 1

$$\frac{df}{dx} = \frac{-2e^{-2x}}{1+e^{-2x}},\ \frac{\partial g}{\partial y} = 2e^{2y} + 6xye^{3xy^2} \tag{41}$$

$\square$

Problem 2

$$\frac{\partial f}{\partial v} = \frac{\partial x}{\partial v}y + x\frac{\partial y}{\partial v} = -\left[\sin\left(u+v\right)\sin\left(u-v\right)\right] - \left[\cos\left(u+v\right)\cos\left(u-v\right)\right] \tag{42}$$

$$= -\cos\left(2v\right) \tag{43}$$

$\square$

Problem 3

$$\int_5^{10}\frac{2}{x-3}dx = \int_5^{10}\frac{2}{x-3}d\left(x-3\right) = 2\int_5^{10}d\left[\ln\left(x-3\right)\right] = 2\ln\left(3.5\right) \tag{44}$$

$\square$

Problem 4

$$\nabla E = \begin{pmatrix} 2ue^{2v} + 4v\left(u-1\right)e^{v-u} - 4v^2e^{-2u} \\ 2u^2e^{2v} - 4u\left(v+1\right)e^{v-u} + 8ve^{-2u} \end{pmatrix} \tag{45}$$

$$\nabla^2 E = \begin{pmatrix} 2e^{2v} + 4v\left(2-u\right)e^{v-u} + 8v^2e^{-2u} & 4ue^{2v} + 4\left(v+1\right)\left(2-u\right)e^{v-u} - 8ve^{-2u} \\ 4ue^{2v} + 4\left(u-1\right)\left(v+1\right)e^{v-u} - 16ve^{2u} & 4u^2e^{2v} - 4u\left(v+2\right)e^{v-u} + 8e^{-2u} \end{pmatrix} \tag{46}$$

□

Problem 5

$$\begin{aligned} T\left(u,\ v\right) =& E\left(1,\ 1\right) + \left(u-1\right)E_u\left(1,\ 1\right) + \left(v-1\right)E_y\left(1,\ 1\right) \qquad (47) \\ &+ \frac{1}{2!}\left[\left(u-1\right)^2 E_{uu}\left(1,\ 1\right) + 2\left(u-1\right)\left(v-1\right)E_{uv}\left(1,\ 1\right) + \left(v-1\right)^2 E_{vv}\left(1,\ 1\right)\right] + \cdots \end{aligned}$$

□

Problem 6

$$Ae^{\alpha} = 2Be^{-2\alpha} \Rightarrow \alpha = -\frac{1}{3}\ln\left(\frac{A}{2B}\right) \tag{48}$$

□

Problem 7

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^T A\mathbf{w} = \left[\frac{\partial}{\partial w_k}\left(\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}w_iw_j\right)\right]_k \tag{49}$$

$$= \left[\sum_{j=1}^{n} a_{kj}w_j\right]_k + \left[\sum_{i=1}^{n} a_{ik}w_i\right]_k \tag{50}$$

$$= A^T\mathbf{w} + A\mathbf{w} = 2A\mathbf{w} \tag{51}$$

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{b}^T\mathbf{w} = \left[\frac{\partial}{\partial w_k}\left(\sum_{i=1}^{n} b_iw_i\right)\right]_k = \left[b_k\right]_k = \mathbf{b} \tag{52}$$

So $\nabla E\left(\mathbf{w}\right) = A\mathbf{w} + \mathbf{b}$. Also,

$$\frac{\partial}{\partial \mathbf{w}^T}A\mathbf{w} = \left[\left(\frac{\partial}{\partial w_i}\left[\sum_{k=1}^{n} a_{jk}w_k\right]_j\right)^T\right]_i = \left[a_{ji}\right]_{ji} = A \tag{53}$$

Hence, $\nabla^2 E = A$.

$\square$

Problem 8

For $\nabla E(\mathbf{w}) = \mathbf{0} \Rightarrow \mathbf{w} = -A^{-1}\mathbf{b}$. Hence $\operatorname{argmin}_{\mathbf{w}}(E(\mathbf{w})) = -A^{-1}\mathbf{b}$.

$\square$

Problem 9

Consider $\nabla \left[\frac{1}{2}\left(w_1^2 + 2w_2^2 + 3w_3^2\right) - \lambda\left(w_1 + w_2 + w_3 - 11\right)\right] = 0$,

$$\left.\begin{aligned} w_1 - \lambda = 0 \\ 2w_2 - \lambda = 0 \\ 3w_3 - \lambda = 0 \\ w_1 + w_2 + w_3 - 11 = 0 \end{aligned}\right\} \Rightarrow w_1 = 6,\ w_2 = 3,\ w_3 = 2 \tag{54}$$

so $\min \left[\frac{1}{2}\left(w_1^2 + 2w_2^2 + 3w_3^2\right)\right] = 33$.

$\square$

Problem 10

Claim $\nabla E(\mathbf{w}) = -\boldsymbol{\lambda}^T \mathrm{A}$ for some vector $\boldsymbol{\lambda}$, if not, assume $\nabla E(\mathbf{w}) = -\boldsymbol{\lambda}^T \mathrm{A} + \mathbf{u}^T$ and consider

$$E(\mathbf{w} - \eta \cdot \mathbf{u}) \sim E(\mathbf{w}) - \eta \nabla E(\mathbf{w})\,\mathbf{u} = E(\mathbf{w}) - \eta\left(\mathbf{u}^T\mathbf{u} - \boldsymbol{\lambda}^T \mathrm{A}\mathbf{u}\right) = E(\mathbf{w}) - \eta\,\|\mathbf{u}\|^2 \tag{55}$$

for some small $\eta > 0$. Then there must exists $\mathbf{u}$ such that $E(\mathbf{w}) > E(\mathbf{w} - \eta \cdot \mathbf{u})$, which is a contradiction.

$\square$

Reference

[1] Lecture Notes by Hsuan-Tien LIN, Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan.