CS 189 HW8 Write-Up

Qingyang Zhao

TOTAL POINTS

79.58 / 90

QUESTION 1

1 Getting Started 10 / 10

✓ + **10 pts** **Correct**

+ **0 pts** Blank

QUESTION 2

Gaussian Classification 30 pts

2.1 a 10 / 10

✓ + **10 pts** **Correct**

+ **8 pts** Mostly Correct (small computation mistake)

+ **7 pts** decision rule incorrect or missing

+ **5 pts** Halfway

+ **3 pts** First step

+ **0 pts** Incorrect

2.2 b 10 / 10

✓ + **10 pts** **Correct**

+ **8 pts** mostly correct

+ **5 pts** get part of the misclassified probability

+ **0 pts** Incorrect

2.3 c 10 / 10

✓ + **10 pts** **Correct**

+ **6 pts** Partial

+ **0 pts** Incorrect/Missing

QUESTION 3

3 Multiple Choice Questions 8.58 / 10

+ **1.43 pts** Part (a) correct

✓ + **1.43 pts** **Part (b) correct**

✓ + **1.43 pts** **Part (c) correct**

✓ + **1.43 pts** **Part (d) correct**

✓ + **1.43 pts** **Part (e) correct**

✓ + **1.43 pts** **Part (f) correct**

✓ + **1.43 pts** **Part (g) correct**

+ **0 pts** Blank

QUESTION 4

4 Short Answer Questions 5 / 10

+ **10 pts** (A) and (B) correct

✓ + **5 pts** **only (A) correct**

+ **5 pts** only (B) correct

+ **0 pts** Both incorrect

QUESTION 5

5 Parameter Estimation 6 / 10

+ **10 pts** Correct

+ **8 pts** Mostly Correct

✓ + **6 pts** **part b correct or significant errors**

+ **2 pts** part a correct

+ **0 pts** Incorrect

QUESTION 6

6 Linear Regression 10 / 10

✓ + **10 pts** **all correct**

+ **5 pts** only one is correct

+ **0 pts** Incorrect

+ **8 pts** mostly correct

QUESTION 7

7 Finding Noisy Low-Rank Matrices 0 / 0

✓ + **0 pts** **--**

QUESTION 8

8 Multi-view Regression with CCA 0 / 0

✓ + **0 pts** **--**

QUESTION 9

9 Your Own Question 10 / 10

✓ + **10 pts** **Own question made**

+ **0 pts** Missing

HO8-1

a) Names — Email Address

① Wan — jhun0324@berkeley.edu

Description of Team: Best Group Ever

How did I work?

Comments:

b) I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

Qingyang Zhao

1 Getting Started 10 / 10

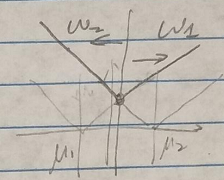
✓ + **10 pts** Correct

+ **0 pts** Blank

$\frac{(x-\mu_1)^2}{2\sigma} =$

HW08 289A (-2)

2. Gaussian Classification.



a) Liklihood: $P(x|w_1)$ $P(x|w_2)$

Decision Boundary: $*$ $P(x|w_1)P(w_1) = P(x|w_2)P(w_2)$

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}}$$

$|x-\mu_1| = |x-\mu_2|$ $\boxed{x = \frac{\mu_1+\mu_2}{2}}$

Decision rule: When $P(w_1|x) > P(w_2|x)$ It's w_1

otherwise It's w_2

Since $P(w_1|x) \propto P(x|w_1)P(w_1)$

$P(w_2|x) \propto P(x|w_2)P(w_2)$

Thus, $P(w_1|x) > P(w_2|x)$

$P(x|w_1) > P(x|w_2)$

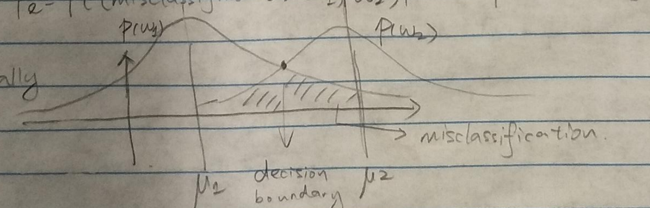
Conclusion, when $\boxed{x > \frac{\mu_1+\mu_2}{2}}$ x belongs to ω_1

Otherwise belongs to w_2.

b) The probability of misclassification rate:

$P_e = P((\text{misclassified as } w_1)|w_2)P(w_2) + P(\text{misclassified as } w_2|w_1)P(w_1)$

Graphically



$$P_e = \frac{1}{2} \cdot \int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma}(x-\mu_2)^2} dx + \frac{1}{2}\int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma}(x-\mu_1)^2} dx$$

$$= \frac{1}{2} \times 2 \int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x}{\sigma}-\frac{\mu_1}{\sigma})^2} dx$$

2.1 a 10 / 10

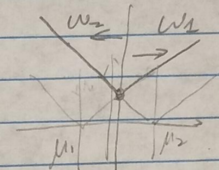
- ✓ + **10 pts** Correct
- + **8 pts** Mostly Correct (small computation mistake)
- + **7 pts** decision rule incorrect or missing
- + **5 pts** Halfway
- + **3 pts** First step
- + **0 pts** Incorrect

$\frac{(x-\mu_1)^2}{2\sigma} =$

HW08 289A (-2)

2. Gaussian Classification.

a) Liklihood: $P(x|w_1)$ $P(x|w_2)$



Decision Boundary: $*$ $P(x|w_1)P(w_1) = P(x|w_2)P(w_2)$

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}}$$

$$|x-\mu_1| = |x-\mu_2| \quad \boxed{x = \frac{\mu_1+\mu_2}{2}}$$

Decision rule: When $P(w_1|x) > P(w_2|x)$ It's w_1

Otherwise It's w_2

Since $P(w_1|x) \propto P(x|w_1)P(w_1)$

$P(w_2|x) \propto P(x|w_2)P(w_2)$

Thus, $P(w_1|x) > P(w_2|x)$

$P(x|w_1) > P(x|w_2)$

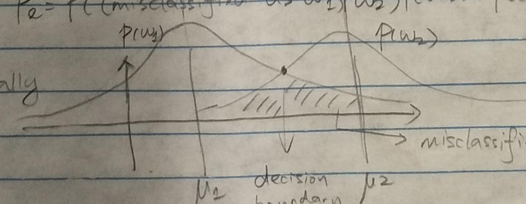
Conclusion, when $\boxed{x > \frac{\mu_1+\mu_2}{2}}$ x belongs to w_1

Otherwise belongs to w_2.

b) The probability of misclassification rate:

$P_e = P((\text{misclassified as } w_1)|w_2)P(w_2) + P(\text{misclassified as } w_2 | w_1)P(w_1)$

Graphically



$$P_e = \frac{1}{2} \cdot \int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma}(x-\mu_2)^2} dx + \frac{1}{2}\int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma}(x-\mu_1)^2} dx$$

$$= \frac{1}{2} \times 2 \int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x}{\sigma}-\frac{\mu_1}{\sigma})^2} dx$$

$$\frac{1}{2}\times 2\int_{\frac{\mu_1+\mu_2}{2}}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu_1)^2}{2\sigma^2}}dx \qquad \text{set } z=\frac{x-\mu_1}{\sigma},\ x=\sigma z+\mu_1$$

$$=\int_{\frac{\mu_1+\mu_2}{2}}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}z^2}d\sigma z+\mu_1 \qquad x\in\left(\frac{\mu_1+\mu_2}{2},+\infty\right) \quad z=\frac{x-\mu_1}{\sigma}\in\left(\frac{\mu_2-\mu_1}{2\sigma},+\infty\right)$$

$$\boxed{=\int_{\frac{\mu_2-\mu_1}{2\sigma}}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz}$$

c) $a\to\infty$

$$\lim_{a\to\infty}P_e=\lim_{a\to\infty}\frac{1}{\sqrt{2\pi}}\int_{a}^{\infty}e^{-z^2/2}dz=0$$

2.2 b 10 / 10

✓ + **10 pts** Correct

+ **8 pts** mostly correct

+ **5 pts** get part of the misclassified probability

+ **0 pts** Incorrect

$$\frac{1}{2} \times 2 \int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} dx \qquad \text{set } z = \frac{x-\mu_1}{\sigma}, \; x = \sigma z + \mu_1$$

$$= \int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2} d\sigma z + \mu_1 \qquad x \in \left(\frac{\mu_1+\mu_2}{2}, +\infty\right) \qquad z = \frac{x-\mu_1}{\sigma} \in \left(\frac{\mu_2-\mu_1}{2\sigma}, +\infty\right)$$

$$= \boxed{\int_{\frac{\mu_2-\mu_1}{2\sigma}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz}$$

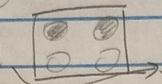
c) $a \to \infty$

$$\lim_{a\to\infty} P_e = \lim_{a\to\infty} \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{-z^2/2} dz = 0$$

2.3 C 10 / 10

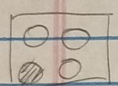
- ✓ + **10 pts** Correct
- + **6 pts** Partial
- + **0 pts** Incorrect/Missing

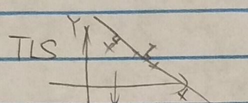
HW08 (-3)

a) $\lambda \to \uparrow$ $\omega \to \downarrow$  which means a smaller variance (small model complexity)

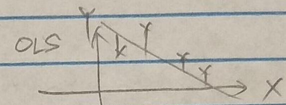
Gaussian Prior with smaller variance

Uniform Prior with smaller variance $Var(Uniform) = \frac{1}{12}(b-a)^2$ range$\downarrow$ var$\downarrow$

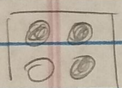
b) 

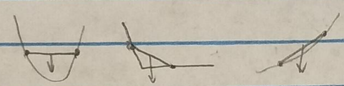


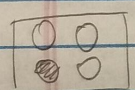
Because allow noise in X, Y

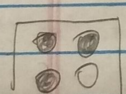


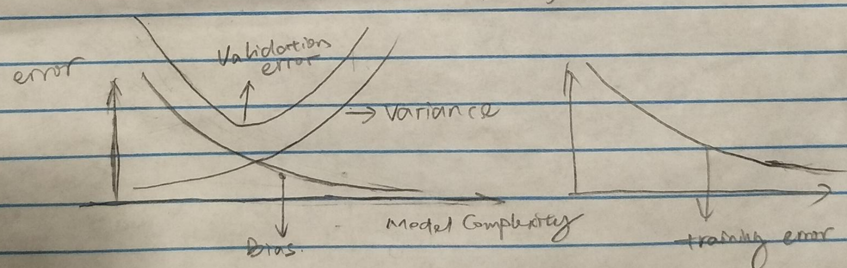
Only Allow noise in Y

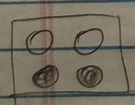
c) 

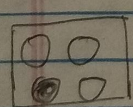


d)  (x,y) negative related, not in a line. $\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$

e) 



f) 

g) 

3 Multiple Choice Questions 8.58 / 10

+ **1.43 pts** Part (a) correct
- ✓ + **1.43 pts** Part (b) correct
- ✓ + **1.43 pts** Part (c) correct
- ✓ + **1.43 pts** Part (d) correct
- ✓ + **1.43 pts** Part (e) correct
- ✓ + **1.43 pts** Part (f) correct
- ✓ + **1.43 pts** Part (g) correct

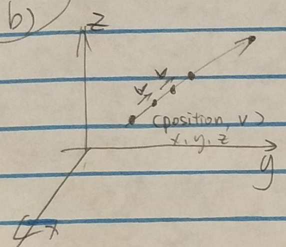
+ **0 pts** Blank

H08-4

a) $a = \rho\sigma_1\sigma_2 \quad \sigma_1 = 2 \quad \sigma_2 = 1$

$\rho \in [-1, 1] \quad a \in [-2, 2]$

b)



$$\begin{bmatrix} x_p^{(1)} & y_p^{(1)} & z_p^{(1)} & x_v^{(1)} & y_v^{(1)} & z_v^{(1)} \\ \vdots & & & & & \vdots \\ x_p^{(n)} & y_p^{(n)} & z_p^{(n)} & & & z_v^{(n)} \end{bmatrix}$$

1 component for position, 1 component for velocity

4 Short Answer Questions 5 / 10

+ **10 pts** (A) and (B) correct

✓ + **5 pts** **only (A) correct**

+ **5 pts** only (B) correct

+ **0 pts** Both incorrect

HW08 -5

(a) $\ln p(x_1, x_2 \cdots x_n|\lambda)$

$\ln \prod_{i=1}^{n} p(x_i|\lambda)$

$= \sum_{i=1}^{n} \ln p(x_i|\lambda) = \sum_{i=1}^{n} \ln \frac{\lambda^{x} e^{-\lambda}}{x!} = \sum_{i=1}^{n} (x \ln\lambda - \lambda - \ln x_i!)$

(b) ① negative log liklihood is a convex function.

$$\frac{\partial -\ln p(x_1, x_2 \cdots x_n|\lambda)}{\partial \lambda} = -\sum_{i=1}^{n} \left(\frac{x_i}{\lambda} - 1\right)$$

$$\frac{\partial^2 -\ln p(x_1, x_2 \cdots x_n|\lambda)}{\partial \lambda^2} = \sum_{i=1}^{n} \frac{x_i}{\lambda^2}$$

Since $x_i > 0$ $\sum_{i=1}^{n} \frac{x_i}{\lambda^2} > 0$

Then negative log liklihood is a convex function.

② Maximum Logliklihood Estimate:

$$\frac{\partial -\ln}{\partial \lambda} = 0 \quad \lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$

c) MAP: $\ln[p(x_1 x_2 \cdots|\lambda) \cdot p(\lambda)]$

$= \sum_{i=1}^{n} (x \ln\lambda - \lambda - \ln x_i!) + \ln \alpha e^{-\alpha\lambda}$

$= \sum_{i=1}^{n} (x_i \ln\lambda - \lambda - \ln x_i!) + \ln\alpha - \alpha\lambda$

$$\frac{\partial l_{MAP}}{\partial \lambda} = \sum_{i=1}^{n} \left(\frac{x_i}{\lambda} - 1\right) - \alpha = 0$$

$$\lambda = \frac{n + \alpha}{\sum_{i=1}^{n} x_i}$$

$\boxed{\text{They both give the true } \lambda^*}$

As $n \to \infty$ $\lim_{n\to\infty} \lambda_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{n}{nE[X]} = \lambda^*$ $\quad \lim_{n\to\infty} \lambda_{MAP} = \frac{n+\alpha}{\sum_{i=1}^{n} x_i} = \frac{n+\alpha}{nE[X]} = \frac{n+\alpha}{n\lambda^*} = \lambda^*$

5 Parameter Estimation 6 / 10

+ **10 pts** Correct
+ **8 pts** Mostly Correct
✓ + **6 pts** **part b correct or significant errors**
+ **2 pts** part a correct
+ **0 pts** Incorrect

H08-6 Linear Regression

a)

$X' = \begin{bmatrix} X \\ cI \end{bmatrix}$

$w' = (X'^T X')^{-1} X'^T y'$

$\left(\begin{bmatrix} X^T & cI \end{bmatrix} \begin{bmatrix} X \\ cI \end{bmatrix} \right)^{-1} \begin{bmatrix} X^T & cI \end{bmatrix} \begin{bmatrix} y \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$= (X^T X + c^2 I)^{-1} X^T y$

where $c^2 = \lambda$ $c = \sqrt{\lambda}$

b) $\min_x f(x) = \min_x \frac{1}{2} \|X' w' - y'\|_2^2$

when $X'^T X'$ is positive definite and using a step size

$\Gamma = \dfrac{2}{\lambda_{min}(X'^T X') + \lambda_{max}(X'^T X')}$

$X'^T X' = X^T X + c^2 I = \lambda_{min}(X'^T X') = m + \lambda$

$\lambda_{max}(X'^T X') = M + \lambda$

$\gamma = \dfrac{2}{2\lambda + m + M}$

6 Linear Regression 10 / 10

- ✓ + **10 pts** **all correct**
- \+ **5 pts** only one is correct
- \+ **0 pts** Incorrect
- \+ **8 pts** mostly correct

7. M^*

a) $P(Y|M) \sim N(y_{ij} - m_{ij}, 1)$

$\prod_{i,j} P(Y_{ij}|M) = \prod_{ij} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_{ij}-m_{ij})^2}{2}}$

$= \frac{1}{\sqrt{2\pi}} e^{-\sum_{i,j}(y_{ij} - m_{ij})^2}$

$\ln P(Y|M) = -\frac{\sum_{i,j}(y_{ij}-m_{ij})^2}{2} + \text{const}$

$\underset{M}{\text{minimize}} \ \|Y-M\|_F$

b) $\hat{M} = \underset{M: \text{rank}(M) \le k}{\arg\min} \|Y - M\|_F^2$ (N)

$\underset{M}{\min} \|Y - M\|_F^2$ $\iff$ $\underset{N}{\text{minimize}} \|N\|_F^2$

$Mx = 0$ $\qquad [Y-N]x = 0$

Eckart-Young's Theorem, the solution of left form of N^* is

$N^* = \sigma_d u_d v_d^T$ If $Y = U_Y D_Y V_Y^T = \begin{bmatrix} u_1 & \cdots & u_d \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_d \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_d^T \end{bmatrix}$

Thus $M^* = \sum_{i=1}^{d-1} \sigma_i u_i v_i^T$

c) Also According to Eckart-Young's Theorem,

$N^* = \sum_{i=k+1}^{d} \sigma_i u_i v_i^T$

$M^* = \sum_{i=1}^{k} \sigma_k u_k v_k^T$

7 Finding Noisy Low-Rank Matrices 0 / 0

✓ + **0 pts** --

HW08-8

a) $E[\hat{X}\hat{Q}^T] = E[U^T \Sigma_{XX}^{-\frac{1}{2}} X Q \Sigma_{QQ}^{-\frac{1}{2}} V]$ $\qquad \Sigma_{XX}^{-\frac{1}{2}} \Sigma_{XQ} \Sigma_{QQ}^{-\frac{1}{2}} = U \Lambda V^T$

$= E[U^T U \Lambda V^T V]$

$= E[\Lambda] = \Lambda$

$\Sigma_{\hat{X}\hat{X}} = E[\hat{X}\hat{X}^T] = E[U^T \Sigma_{XX}^{-\frac{1}{2}} X X^T \Sigma_{XX}^{-\frac{1}{2}} U]$

$= U^T \Sigma_{XX}^{-\frac{1}{2}} E[XX^T] \Sigma_{XX}^{-\frac{1}{2}} U$

$= U^T \Sigma_{XX}^{-\frac{1}{2}} \Sigma_{XX} \Sigma_{XX}^{-\frac{1}{2}} U = I$

$\Sigma_{\hat{Q}\hat{Q}} = E[\hat{Q}\hat{Q}^T] = E[V^T \Sigma_{QQ}^{-\frac{1}{2}} Q Q^T \Sigma_{QQ}^{-\frac{1}{2}} V]$

$= V^T \Sigma_{QQ}^{-\frac{1}{2}} \Sigma_{QQ} \Sigma_{QQ}^{-\frac{1}{2}} V$

$= I$

b) $E[(Y - w^T\hat{X})^2] = E[Y^2] + w^T E[\hat{X}\hat{X}^T] w - 2E[Y w^T \hat{X}]$

$= E[Y^2] + \|w\|_2^2 - 2w^T E[Y\hat{X}]$

c) $\hat{w} = \underset{w \in \mathbb{R}^d}{\operatorname{argmin}} E[Y^2] + \|w\|_2^2 - 2w^T E[Y\hat{X}] + \|w\|^2_{CCA}$

$= \underset{w \in \mathbb{R}^d}{\operatorname{argmin}} E[Y^2] + \|w\|_2^2 - 2w^T E[Y\hat{X}] + \sum_{i=1}^{d} \frac{1-\lambda_i}{\lambda_i} (w_i)^2 = f(w)$

$\frac{\partial f(w)}{\partial w_i} = 0 \qquad 2w_i - 2E[Y\hat{X}_i] + 2\sum_{i=1}^{d} \frac{1-\lambda_i}{\lambda_i} w_i = 0$

$w_i = \frac{E[Y\hat{X}_i]}{1 + \frac{1-\lambda_i}{\lambda_i}} = \lambda_i E[Y\hat{X}_i]$

d) $E[\|\tilde{\omega} - \hat{\omega}\|_2^2] \leq \sum_{i=1}^{d} \frac{\lambda_i^2}{n}$

↗ variance

$$E[\|\tilde{\omega} - \hat{\omega}\|_2^2] = E[\|\tilde{\omega} - E[\tilde{\omega}]\|_2^2]$$

$$= E[(\tilde{\omega} - E[\tilde{\omega}])^T(\tilde{\omega} - E[\tilde{\omega}])]$$

$$= E[\tilde{\omega}^T\tilde{\omega} - E[\tilde{\omega}]^T\tilde{\omega} + E[\tilde{\omega}]^T E[\tilde{\omega}]]$$

$$= E[\tilde{\omega}^T\tilde{\omega}] - E[\tilde{\omega}]^T E[\tilde{\omega}]$$

$$= \sum_{i=1}^{d} E[\tilde{\omega}_i^2] - E[\tilde{\omega}_i]^2$$

$u_j^T u_j$

$$E[\tilde{\omega}_i^2] = E\left[\lambda_i^2 \frac{1}{n^2}\left(\sum_{j=1}^{n} y^j (\hat{x}^j)_i\right)^2\right]$$

$$= \frac{\lambda_i^2}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} E[y^j \hat{x}_i^j y^k \hat{x}_i^k]$$

$$\hat{x}^j = \begin{bmatrix} u_1^T \\ \vdots \\ u_d^T \end{bmatrix} \Sigma_{xx}^{-\frac{1}{2}} x^j$$

$$\hat{x}_i^j = u_i^T \Sigma_{xx}^{-\frac{1}{2}} x^j$$

$$\hat{x}_i^j = u_i^T \Sigma_{xx}^{-\frac{1}{2}} x^j$$

$$\hat{x}_i^j \cdot \hat{x}_i^k = \left(u_j^T \Sigma_{xx}^{-\frac{1}{2}} x_j\right) \cdot \left(u_k^T \cdot \Sigma_{xx}^{-\frac{1}{2}} x_k\right)$$

$$E[\tilde{\omega}_i^2] = \frac{\lambda_i^2}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} E[u_j^T \Sigma_{xx}^{-\frac{1}{2}} x_j \cdot x_k^T \Sigma_{xx}^{-\frac{1}{2}} u_k \cdot y^j \cdot y^k]$$

$$= \frac{\lambda_i^2}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} u_j^T \Sigma_{xx}^{-\frac{1}{2}} E[x_j \cdot x_k^T \cdot y^j \cdot y^k] \Sigma_{xx}^{-\frac{1}{2}} u_k$$

$$\leq \frac{\lambda_i^2}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} u_j^T \cdot \Sigma_{xx}^{-\frac{1}{2}} \Sigma_{xx} \Sigma_{xx}^{-\frac{1}{2}} u_k$$

$$u_j^T u_k = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

$$\boxed{E[x_j x_k^T \cdot y^j \cdot y^k] \leq E[x_j x_k^T] \text{ for } y^2 \leq 1}$$

Since $E[\tilde{\omega}_i]^T E[\tilde{\omega}_i] > 0$

Then, $E[\|\tilde{\omega} - \hat{\omega}\|_2^2] = E[\tilde{\omega}^T\tilde{\omega}] - \overset{0}{E[\tilde{\omega}]^T E[\tilde{\omega}]}$

$$\leq \sum_{i=1}^{d} \frac{\lambda_i^2}{n}$$

8 Multi-view Regression with CCA 0 / 0

✓ + **0 pts** --

H8 My Own Question

Bias-Variance Decomp in training / test data.

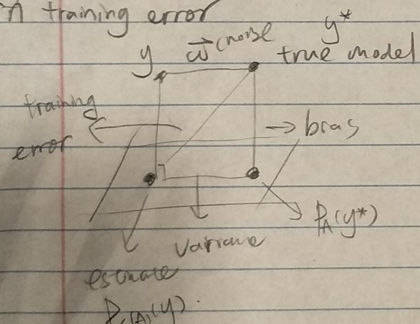
$E[Y] = E[f(x)+N] = f(x) + E[N] = f(x)$

$\mathcal{E}(x;h) = E[(h(x;D)-Y)^2] = E[h(x;D)^2] + E[Y^2] - 2E[\underbrace{h(x;D)}_{\text{from test data}} \cdot \underbrace{Y}_{\text{training data, independent}}]$

$= E[h(x;D)^2] + E[Y^2] - 2E[h(x;D)]E[Y]$

$= \underbrace{(E[h(x;D)] - E[Y])^2}_{\text{bias}} + \underbrace{V[h(x;D)]}_{\text{var}} + \underbrace{V[N]}_{\text{irre}}$

bias-var. $\rightarrow$ Linear Model in training error



$\underbrace{E[\|y^* - P_A(y^*+w)\|_2^2]}_{\text{training error}} = \underbrace{\|y^* - Ax^*\|_2^2}_{\text{bias}} + \underbrace{E[\|Ax^* - A\hat{x}\|_2^2]}_{\text{var}}$

doesn't contain irreducible error and noise in training data related to variance.

9 Your Own Question 10 / 10

✓ + **10 pts** **Own question made**

+ **0 pts** Missing