Daniel Suo
ELE 535
HW1

Q1 a) $\hat{y}(x) = w^T x + b$

$$= (\hat{\mu}_0 - \hat{\mu}_1)^T x + (\hat{\mu}_1 - \hat{\mu}_0)^T \frac{\hat{\mu}_1 + \hat{\mu}_0}{2}$$

$$0 = \sum_{i=0}^{n-1} (\hat{\mu}_{0i} - \hat{\mu}_{1i}) x_i + \frac{\hat{\mu}_{1i}^2 - \mu_{0i}^2}{2}$$

For each i:
$$0 = 2\hat{\mu}_{0i}x_i - 2\hat{\mu}_{1i}x_i + \mu_{1i}^2 - \mu_{0i}^2$$

$$0 = x_i^2 - x_i^2 + 2\hat{\mu}_{0i}x_i - 2\hat{\mu}_{1i}x_i + \hat{\mu}_{1i}^2 - \hat{\mu}_{0i}^2$$

$$0 = (x_i - \hat{\mu}_{1i})^2 - (x_i - \hat{\mu}_{0i})^2$$

Vector sum:
$$0 = \sum_{i=0}^{n-1} (x_i - \hat{\mu}_{1i})^2 - (x_i \hat{\mu}_{0i})^2$$

$$= \|x - \hat{\mu}_1\|_2^2 - \|x - \hat{\mu}_0\|_2^2$$

Inequality:
$$\|x - \hat{\mu}_0\|_2 > \|x - \hat{\mu}_1\|_2$$

$\therefore$ $\hat{y}(x) = w^T x + b < 0$ is the nearest centroid classifier

Q1 b) Observe that $\hat{\mu} = \frac{\hat{\mu}_1 + \hat{\mu}_0}{2}$ because there are an equal # of examples in both classes.

$$\langle \hat{\mu}_0 - \hat{\mu}_1, x - \hat{\mu} \rangle = \sum_{i=0}^{n-1} (\hat{\mu}_{0i} - \hat{\mu}_{1i}) x_i - (\hat{\mu}_{0i} - \hat{\mu}_{1i}) \left(\frac{\hat{\mu}_{0i} + \hat{\mu}_{1i}}{2} \right)$$

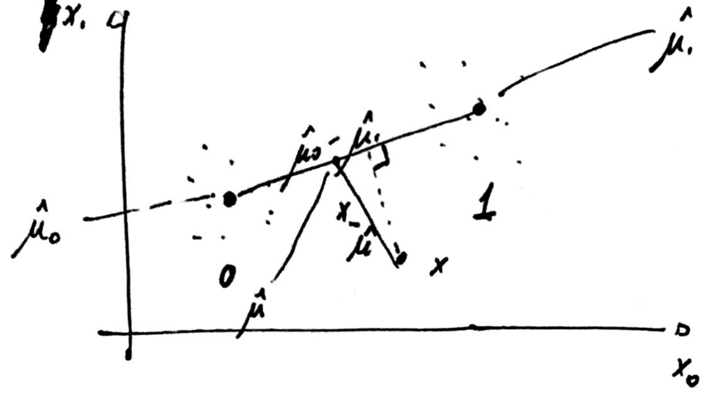
$$= \sum_{i=0}^{n-1} \frac{x_i^2}{2} - \frac{x_i^2}{2} + \hat{\mu}_{0i} x_i - \hat{\mu}_{1i} x_i - \frac{\hat{\mu}_{0i}^2}{2} + \frac{\hat{\mu}_{1i}^2}{2} < 0$$

$$\|x - \hat{\mu}_1\|_2 < \|x - \hat{\mu}_0\|_2$$

$\therefore$ $\hat{y}(x) = \begin{cases} 1, & \text{if } \langle \hat{\mu}_0 - \hat{\mu}_1, x - \hat{\mu} \rangle < 0 \\ 0 & \text{otherwise} \end{cases}$ is nearest centroid classifier

Daniel Suo
ELE 535
HW1

Q1 c)



The dot product of $\hat{\mu}_0 - \hat{\mu}_1$ and $x - \hat{\mu}$ represents the projection of the input x onto $\hat{\mu}_0 - \hat{\mu}_1$ at $\hat{\mu}$. If x is closer to $\hat{\mu}_0$ along $\hat{\mu}_0 - \hat{\mu}_1$ (i.e., $\hat{\mu}$ as threshold), then classify x as 0. Classify x as 1 otherwise.

Q1 d) If we subtract $\hat{\mu}$ from all training data:

$$\hat{\mu}_0' = \hat{\mu}_0 - \hat{\mu}$$
$$\hat{\mu}_1' = \hat{\mu}_1 - \hat{\mu}$$
$$\hat{\mu}' = \hat{\mu} - \hat{\mu} = 0$$

We also must normalize any input x

$$\hat{y}'(x) = \begin{cases} 1, & \text{if } \langle \hat{\mu}_0' - \hat{\mu}_1', x - \hat{\mu} \rangle \\ 0, & \text{otherwise} \end{cases}$$

This result makes sense b/c

$$\hat{\mu}_0' - \hat{\mu}_1' = \hat{\mu}_0 - \hat{\mu}_1$$

Daniel Suo
ELE 535
HW 1

Q2 a) $R(X)$ is the image of X,
which is a subspace of its range, $\mathbb{R}^n$.

i) $R(X)$ contains the 0 vector:

$$w = 0 \in \mathbb{R}^m \Rightarrow z = X0 = 0 \in \mathbb{R}^n$$

ii) $R(X)$ closed under linear combination

$$w = ax + by \ ; \ x, y \in \mathbb{R}^n$$

$$z = X(ax + by) = aXx + bXy \in \mathbb{R}^n$$

iii) $R(X) \subset \mathbb{R}^n$

Trivially true by property of matrix multiplication

$$R(X) = \{ z \in \mathbb{R}^n : z = Xw, \ w \in \mathbb{R}^m \}$$

all $z \in \mathbb{R}^n$

Q2 b) $N(X)$ is the kernel of X,
which is a subspace of its domain, $\mathbb{R}^m$

i) $N(X)$ contains the 0 vector

$$a = 0 \in \mathbb{R}^m \rightarrow Xa = 0 \in N(X)$$

ii) $N(X)$ closed under linear combination

$$a = mx + by \ ; \ x, y \in N(X)$$

$$X(mx + by) = mXx + bXy$$

$$= m0 + b0 = 0$$

iii) $N(X) \subset \mathbb{R}^m$

By defn, all elements of $N(X)$ in $\mathbb{R}^m$

Daniel Suo
ELE 535
HW1

Q3a) $N_1 \cap N_2$ is the intersection of two kernels of matrices w/ same domain $\mathbb{R}^m$

i) $0 \in N_1 \cap N_2$

Trivially true since 0 vector is in both null spaces, N_1 and N_2

ii) $x_1, x_2 \in N_1 \cap N_2$

x_1 and x_2 are in N_1 and N_2, so

$ax_1 + bx_2$ is in both kernel of A_1 and A_2

per Q2b. Thus $ax_1 + bx_2 \in N_1 \cap N_2$

iii) $N_1 \cap N_2 \in \mathbb{R}^m$

$N_1 \in \mathbb{R}^m$ and $N_2 \in \mathbb{R}^m$ by Q2b.

Thus $N_1 \cap N_2 \in \mathbb{R}^m$

~~Q3b)~~ We can represent $N_1 \cap N_2$ as all solutions to the matrix equation

Equation $\longrightarrow$ $\|A_1 x\| + \|A_2 x\| = 0$.

This ensures that x is in kernel for both A_1 and A_2 and thus in $N_1 \cap N_2$.

Daniel Su
ELE 535
HW1

Q3b) $R_1 + R_2$ ~~[illegible]~~

i) $R_1 + R_2 \ni 0$

$x_1 = 0 \in \mathbb{R}^{m_1}$, $x_2 = 0 \in \mathbb{R}^{m_2}$

$y_1 = A_1 x_1 = 0$

$y_2 = A_2 x_2 = 0$

$y_1 + y_2 = 0 \in R_1 + R_2$

ii) closed under linear combo

let $a = y_1 + y_2$ ~~[illegible]~~ $\in R_1 + R_2$

$b = y_3 + y_4$, ~~[illegible]~~ $\in R_1 + R_2$

$y_1, y_3 \in R_1$

$y_2, y_4 \in R_2$

$c_1 a + c_2 b \in R_1 + R_2$

$c_1 (y_1 + y_2) + c_2 (y_3 + y_4)$

$c_1 y_1 + c_2 y_3 \in R_1$ under linearity

$c_1 y_2 + c_2 y_4 \in R_2$ under linearity

$(c_1 y_1 + c_2 y_3) + (c_1 y_2 + c_2 y_4) \in R_1 + R_2$

iii) $R_1 + R_2 \in \mathbb{R}^n$ by defn of addition of vectors, any $y_1, y_2 \in \mathbb{R}_n$, $y_1 + y_2 \in \mathbb{R}_n$

Equation $\rightarrow$ $R(x_1, x_2) = R_1(x_1) + R_2(x_2) = A_1 x_1 + A_2 x_2$

$x_1 \in \mathbb{R}^{m_1}$, $x_2 \in \mathbb{R}^{m_2}$

Daniel Suo
ELE 535
HW1

Q4). $A, B \in \mathbb{R}^{m \times n}$

$A^T \in \mathbb{R}^{n \times m}$

$A^T \cdot B \in \mathbb{R}^{n \times n}$

Let $C = A^T \cdot B$

$$C_{ij} = \sum_{k=0}^{m-1} A^T_{ik} B_{kj} = \sum_{k=0}^{m-1} A_{ki} B_{kj}$$

$$\text{trace}(C) = \sum_{i=0}^{n-1} C_{ii} = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} A_{ji} B_{ji}$$

Q5) for all $s \in S$, $s_{ij} = s_{ji}$

for all $a \in A$, $a_{ij} = -a_{ji}$

$a_{ii} = 0$

i) S is subspace of $\mathbb{R}^{n \times n}$

- 0 vector ($n \times n$ matrix) is symmetric
- S closed under linear combo.

$C = as_1 + bs_2$ $\quad c_{ij} = as_{1ij} + bs_{2ij}$

$c_{ji} = as_{1ji} + bs_{2ji}$

$c_{ij} = c_{ji}$

- $S \subset \mathbb{R}^{n \times n}$ by construction

ii) A is subspace of $\mathbb{R}^n$

- 0 vector ($n \times n$) is anti symmetric
- A closed under linear combo

$c = as_1 + bs_2$ $\quad c_{ij} = as_{1ij} + bs_{2ij}$

$c_{ji} = as_{1ji} + bs_{2ji}$

- $A \subset \mathbb{R}^{n \times n}$ by construction $= -(as_{1ij} + bs_{2ij})$

Daniel Suo
ELE 535
HW1

Q5) $S^{\perp} = A$ defined as $\langle S, A \rangle = 0$

$$\langle S, A \rangle = \text{trace}(S^T A)$$

$$= \text{trace}(S A)$$

$$= \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} S_{ij} A_{ij}$$

For non-diagonal entries,

$$s_{ij} a_{ij} = s_{ij}(-a_{ji}) = -s_{ji} a_{ji}$$

For diagonal entries,

$$s_{ii} a_{ii} = s_{ii}(-a_{ii}) = s_{ii} 0 = 0$$

$$\langle S, A \rangle = 0$$

Thus $\mathbb{R}^{n \times n}$ is direct sum of S and A.

by definition of direct sum.