Mid1_Key

September 24, 2018

1 Set A

1.1 Question 1

1. (e) (Unless mentioned the assumptions multiple answers are not considered)
2. $\frac{-c}{b}$
3. **True**
4. (b) vector
5. **True**. (rank can never be greater than number of rows of matrix)
6. **True**
7. (e)
8. **True** (Bayesian gives the minimum error of all)
9. **False** (has both classes)
10. **True**

1.2 Question 2

.1. (a) covariance is sigma squared the identity matrix.

(b) covariace for the two matrices are equal.

Note: 1 mark has been deducted in case studen has written only one of the above condition

2. Samples should be Linearly seperable.
3. (W.T.)X/|W|, where W are the co-efficients of the equation of the plane. X is the point in 3-d space.
4. Linear in nature with an area of ambiguity in the middle.
5. chapter 2 equation 20, Duda-Hart

1.3 Question 3 (Same method for both sets)

$P(hit\_target/height = 165cm) = \sum_{G=(M,F)} P(hit\_target/G) * P(G/height = 165cm)$

(5 marks)
$P(hit\_target/height = 165cm) = \sum_{G=(M,F)} P(hit\_target/G) * \frac{P(height=165cm/G)*P(G)}{\sum_{G=(M,F)} P(height=165cm/G)*P(G)}$
(5 marks)
$P(hit\_target/height = 165cm) = \sum_{G=(M,F)} P(hit\_target/G) * \frac{\mathcal{N}(\mu_G,\sigma_G)*P(G)}{\sum_{G=(M,F)} \mathcal{N}(\mu_G,\sigma_G)*P(G)}$
(2 marks)
Final answer with correct method $(71 - 73\%)$
(3 marks)

1.4 Question 4

Code given below.

1.4.1 Marks Distribution

- General equation and theroy about perceptron algorithm **(3 marks)**
  - Write the update rule with proper eta value
  - Define the imput vectors and gradient properly
- Iteration 1 **(6 marks)**
- Show the value obtained by the dot product of weight vector and imput vector
- Show which all are misclassified
- Update according to the rule
- Write the equation of the line
- Iteration 2 **(6 marks)**
- Show the value obtained by the dot product of weight vector and imput vector
- Show which all are misclassified
- Update according to the rule
- Write the equation of the line

1.5 Question 5

The following is the rough breakdown of the marking rubric. Each part of the question i.e. "a,b,c" carries 5 marks each.

$5(a)$

Subsection	Marks
pseudo code for f()	2
Minimize or maximize objective function	1
Reason	2

$5(b)$

Subsection	Marks
Equation for objective function (have done binary marking)	5

$5(c)$

Subsection	Marks
Gradient Descent Derivation	3
Pseudo code	1
Reason	1

1.5.1 Solution

1.5.2 a.

f(x) = $\begin{cases} 1 & if \;\; \mathbf{w^T x} > 0 \\ -1 & else \end{cases}$

Minimize the objective!

Function is not differentiable so can't be done in this manner

1.5.3 b

$$J = \frac{1}{N}\sum_{i=1}^{N} 1 - \frac{y_i \cdot f(x_i)}{4}$$

$$or$$

$$J = \frac{1}{N}\sum_{i=1}^{N} 4 - y_i \cdot f(x_i)$$

$$where, f(x) = \begin{cases} 2 & if \;\; \mathbf{w^T x} > 0 \\ -2 & else \end{cases}$$

1.5.4 c

Derive gradient descent using Taylor Series as done in the class

$$J = \frac{1}{N}\sum_{i=1}^{N} 1 - y_i \cdot \sigma(\mathbf{w^T x})$$

$$\mathbf{w}(t+1) = w(t) - \eta \frac{\partial J}{\partial w}$$

$$\nabla_w J = \frac{-1}{N}\sum_{i=1}^{N} y_i \cdot (\sigma(w^T x)(1 - \sigma(w^T x))$$

Sigmoid is non convex!
General Mistake: Very Few people did the derivation part!

1.6 Question 6

The following is the rough breakdown of the marking rubric.

Subsection	Marks
Derivation of newton's update	5
Updated gradient descent	1
Dimensionality, Computation	4
Derivation for MSE for Newton's update	5

1.6.1 Solution

Derivation of Newton's update

$$
\begin{aligned}
f(\mathbf{w}^{\mathbf{n+1}}) &= f(\mathbf{w}^n) - \Delta\mathbf{w}^T\nabla f + \frac{1}{2}\Delta\mathbf{w}^T\mathbf{H}\Delta\mathbf{w} + \dots \\
f(\mathbf{w}^{\mathbf{n+1}}) - f(\mathbf{w}^n) &= -\Delta\mathbf{w}^T\nabla f + \frac{1}{2}\Delta\mathbf{w}^T\mathbf{H}\Delta\mathbf{w} \\
0 &= -\nabla f + \frac{2}{2}\mathbf{H}\Delta\mathbf{w} && \text{Differentiating w.r.t } \boldsymbol{\Delta}\mathbf{w} \\
\Delta\mathbf{w} &= \mathbf{H}^{-1}\nabla f
\end{aligned}
$$

Gradient Descent Equation:

$$
\begin{aligned}
\mathbf{w}(t+1) &= \mathbf{w}(t) - \eta\Delta\mathbf{w} \\
\mathbf{w}(t+1) &= \mathbf{w}(t) - \eta\mathbf{H}^{-1}\nabla f
\end{aligned}
$$

Newton's Update for MSE

$$
f = \sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x}_i)^2
$$

Working out for this particular f, we obtain

$$
\begin{aligned}
\nabla f &= \sum_{i=1}^{N} 2(y_i - \mathbf{w}^T\mathbf{x}_i)(-\mathbf{x}_i) \\
\mathbf{H} &= \sum_{i=1}^{N} 2\mathbf{x}_i\mathbf{x}_i^T
\end{aligned}
$$

Alternate Derivation

$$\begin{aligned} f &= (Y - X\mathbf{w})^T(Y - X\mathbf{w}) \\ &= Y^TY - Y^TX\mathbf{w} - \mathbf{w}^TX^TY + \mathbf{w}^TX^TX\mathbf{w} \\ &= Y^TY - 2Y^TX\mathbf{w} + \mathbf{w}^TX^TX\mathbf{w} \end{aligned}$$

$$\begin{aligned} \nabla_{\mathbf{w}} f &= -2Y^TX + 2\mathbf{w}^TX^TX \\ &= -2(Y - X\mathbf{w})^TX \\ \nabla^2_{\mathbf{w}} f &= 2X^TX \\ \mathbf{H} &= 2X^TX \end{aligned}$$

Deriving Pseudoinverse doesn't get you marks, you're asked to derive terms for Newton's update considering MSE as objective.

2 Set B

2.1 1

1. (e) $0 \leq p(x) \leq 1$. (Unless mentioned the assumption multilple answers are not considered)
2. $\frac{-a}{b}$.
3. False $p(x|c) = p'(x)$.
4. (b)
5. False. Counter example. $\det(\alpha I_n) = \alpha$. Take $n \leq \alpha$.
6. False
7. a,b,c,d
8. True
9. False.
10. True.

2.2 2

1. (a) covariance matrice for each of the two classes should be arbitrary
2. Samples should be Linearly seperable.
3. (W.T)X = 0 , where W are the co-efficients of the equation of the plane. X is the point in 3-d space.
4. The dicsion boundry will be linear in nature.
5. chapter 2 equation 20, Duda-Hart

2.3 4

2.4 Question 4

Code given below.

2.4.1 Marks Distribution

- General equation and theroy about perceptron algorithm **(3 marks)**
  - Write the update rule with proper eta value
  - Define the imput vectors and gradient properly
- Iteration 1 **(6 marks)**
- Show the value obtained by the dot product of weight vector and imput vector
- Show which all are misclassified
- Update according to the rule
- Write the equation of the line
- Iteration 2 **(6 marks)**
- Show the value obtained by the dot product of weight vector and imput vector
- Show which all are misclassified
- Update according to the rule
- Write the equation of the line

2.5 5

The following is the rough breakdown of the marking rubric. Each part of the question i.e. "a,b,c" carries 5 marks each.

$5(a)$

Subsection	Marks
pseudo code for f()	2
Minimize or maximize objective function	1
Gradient Descent	2

$5(b)$

Subsection	Marks
Equation to Penalize	2
Gradient Descent Derivation	3

$5(c)$

Subsection	Marks
equation for V_i	2
2 stage algorithm	3

2.5.1 a.

$$f(x) = \mathbf{w}^T\mathbf{x}$$

Minimize the objective.

Gradient Descent:

$$\mathbf{w}(t+1) = w(t) - \eta\frac{\partial J}{\partial w}$$

$$\nabla_w J = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x_i}$$

2.5.2 b

$$\nabla_w J = \frac{1}{N}\sum_{i=1}^{N} v_i \cdot \mathbf{x_i}$$

Derive the gradient descent rule using Taylor Series as done in the class.
General Mistake : Very few did the derivation part

2.5.3 c1

Can be handled using thresholding, or any function that penalises the sample which are away from the line (any monotonous funciton of this fashion are correct!).

2.5.4 c2

Initialize: **w** and **v**
Stage 1:
update w, Gradient Descent:
$\nabla_w J = \frac{-1}{N}\sum_{i=1}^{N} v_i \cdot x_i$
$\mathbf{w(t+1) = w(t) - J\frac{\partial J}{\partial w}}$

Stage 2: update v's
$\mathbf{v}_{i+1} = \mathbf{f}(\mathbf{w}^*, \mathbf{x_i})$
$where\ \mathbf{f}\ is\ a\ monotonous\ function\ that\ tends\ to\ penalises\ the\ outliers$

2.6 6

2.6.1 Rough

Section .	**Marks**
Taylor Series	**2**
Gradient Descent	**1**
Derivation of Optimal Eta	**5**
Optimal Eta	**2**
Argument Based on the Above	**5**

What is optimal η?

$$J(\mathbf{w}^{k+1}) = J(\mathbf{w}^k) + \mathbf{s}^T \nabla J + \frac{1}{2}\mathbf{s}^T \mathbf{H}\mathbf{s}$$
$$\mathbf{w}^{k+1} - \mathbf{w}^k = s = -\eta \nabla J(\mathbf{w}^k)$$

Substituting, we obtain

$$J(\mathbf{w}^{k+1}) = J(\mathbf{w}^k) - \eta \nabla J^T \nabla J + \frac{\eta^2}{2} J^T \mathbf{H} \nabla J$$

For optimal η, we differentiate w.r.t η and equate to zero.

$$-\nabla J^T \nabla J + \eta \nabla J^T \mathbf{H} \nabla J = 0$$
$$\eta = \frac{\nabla J^T \nabla J}{\nabla J^T \mathbf{H} \nabla J}$$

Thus we have our optimal eta, which changes every iteration above.

Marks have been given for people who made convex objective assumption and argued that the slope will decrease near the minimum, also those who gave generalized argument depending on slope without making the above assumption. You'd have lost marks if you haven't connected your reasoning to the equation above, like the question asks.