CSE 512: Midterm Oct. 23

Instructions:

- You have 90 continuous minutes to complete the exam, self-timed.
- You are allowed 1 page (front and back) cheat sheet. The cheat sheet must be scanned (or photographed with high resolution) and submitted along with the exam solutions. The time begins when you flip this page.
- You are also allowed a simple calculator. (You may use MATLAB or Python if you agree to only use the simple calculator functions.)
- You may print the exam and write your solutions or use lyx/latex. If you need extra sheets of paper, please label them carefully as to which question they are answering. Make your final answer clear.
- If you choose to handwrite your solutions, you must make sure that the digital scan / photograph is of high enough quality that we can see everything clearly. Anything we can't read, we will not grade.
- You may not discuss any problem with any other student while the exam submission portal is still open. You may not look for answers on the internet or in any notes outside of your cheatsheet.

Name: Venkata Subba Narasa Bharath Meadam

Student ID: 11267286

Time began: 11:30 AM

Time ended: 01:00 PM

Scoring	
Q 1	______ / 20
Q 2	______ / 20
Q 3	______ / 20
Q 4	______ / 20
Q 5	______ / 20
Total	______ / 100

Q1)

a) FALSE

Generative model learns joint Probability distribution.

Discriminative model learns conditinal Probability.

b) FALSE

Polynomial Regression is an example of Linear Regression, and it is not a straight line.

Multiple Linear Regression is a plane and not a line.

c) TRUE

d) TRUE

e) TRUE

Logistic Regression is a non-linear way of fitting a linear model,

Q2)

a) x[i]

First, we will convert all the data into a single unit (mm in our case)

x[i] → (50mm, 10mm, 100mm, 10mm, 5mm, 1mm)

y → (+1, +1, +1, -1, -1, -1)

A simple condition like:

$y = +1$ iff $x[i] >= 10$

-1 otherwise

This has a success rate of 5/6

This looks like a step-function.

We can use a $\boxed{\text{Generalized-linear model.}}$

$\underline{x[2]}$

$x[2] \to (+1, +1, -1, -1, -1, -1)$

$y \to (+1, +1, +1, -1, -1, -1)$

very high correlation (5)(b)

$y = x[2] \Rightarrow (mx + c) \Rightarrow m = 1, c = 0$

$\boxed{\text{Linear model}}$ ✓

$\underline{x[3]}$

$x[3]$ values look, completely random with respect to y.

$\boxed{\text{Model: None}}$ ✓

b)

$$\theta[i] = \begin{cases} \text{large, +ve} \to \text{correlates with label} \\ \text{small, } \approx 0 \to \text{feature not important} \\ \text{large, -ve} \to \text{inversely correlates with label} \end{cases}$$

$$\theta^* = \begin{bmatrix} -1.2 \\ 5 \\ -0.2 \end{bmatrix}$$

Importance :-

$\theta^*[1] \rightarrow 5$ (most important) → Personal allergies

As seen from above, it has a very high correlation with the output and correspondingly has a higher weight.

$\theta^*[0] \rightarrow -1.2$ [2nd important] → Size of bite

Similar to the above argument, we were able to come up with a nice decision line to split data

$\theta^*[2] \rightarrow -0.2$ [least important] → Hair length

x[3] was completely random w.r.t. y and correspondingly the least weight

Note:- we may not pre-process data, but we need to ensure that all rows for a given feature has the same SI units

Q3) $f(\theta) = \frac{1}{m} \sum_{i=1} \exp(-y_i x_i^T \theta)$

<u>Dimensions:-</u>

+4

$x = (m \times n)$

$y = (m \times 1)$

$\theta = (n \times 1)$

$m \rightarrow$ examples

$n \rightarrow$ features

Gradient, $\frac{\partial f}{\partial \theta} = \frac{\partial}{\partial \theta} \left(\frac{1}{m} \sum_{i=1}^{m} \exp(-y_i x_i^T \theta) \right)$

$= \frac{1}{m} \frac{\partial}{\partial \theta} \left(\sum_{i=1}^{m} e^{-y_i x_i^T \theta} \right)$

$= \frac{1}{m} \sum_{i=1}^{m} e^{-y_i x_i^T \theta} (-y_i x_i^T)$

$$= \frac{-1}{m} \sum_{i=1}^{m} e^{-y_i x_i^T \theta} \left(-y_i x_i^T\right)$$

$$\boxed{\begin{array}{l} \text{Gradient} = \frac{-1}{m} \sum_{i=1}^{m} e^{-y_i x_i^T \theta} \left(-y_i x_i^T\right) \\ \text{dimensions} = (n \times 1) \end{array}}$$

what you have written is 1 x n

✓

Hessian, $\nabla^2 f(\theta)$,

$$\Rightarrow \frac{1}{m} \frac{\partial}{\partial \theta} \left(\sum_{i=1}^{m} e^{-y_i x_i^T \theta} \left(-y_i x_i^T\right) \right)$$

$$\Rightarrow \frac{1}{m} \left(\sum_{i=1}^{m} e^{-y_i x_i^T \theta} \left(y_i x_i^T\right)^2 \right)$$

$$\boxed{\begin{array}{l} \text{Hessian} = \frac{1}{m} \sum_{i=1}^{m} e^{-y_i x_i^T \theta} \left(y_i x_i^T\right)^2 \\ \text{dimensions} = (n \times n) \end{array}}$$

can't do this with vector

-1

b) As Hessian, $\nabla^2 f(\theta) = \frac{1}{m} \sum_{i=1}^{m} e^{-y_i x_i^T \theta} (y_i x_i^T)^2$

is PSD (Positive semi-definite),

f is convex.

If, Hessian is PSD, function is convex.

+1

Q4)

(a) $\text{margin} = \dfrac{1}{\|\theta\|}$

$\|\theta\| = \sqrt{1^2 + 3^2 + 5^2} = \sqrt{1+9+25} = \sqrt{35}$

$\text{margin} = \dfrac{1}{\sqrt{35}} = 0.16$ +5

(b) A point is misclassified, iff

$1 - s_i < 0$

$$1 - s_i = \begin{bmatrix} 1 - 1.1 \\ 1 - 0 \\ 1 - (-0.5) \\ 1 - 0 \\ 1 - (-3.9) \\ 1 - 0 \end{bmatrix} = \begin{bmatrix} -0.1 \\ 1 \\ 1.5 \\ 1 \\ 4.9 \\ 1 \end{bmatrix}$$

Points x_1 is misclassified +5

(c) If λ increases, misclassified points decreases. +5

If λ decreases misclassified points increases

$$\boxed{\begin{aligned} &\lambda\ (\uparrow) \rightarrow \text{misclassified points } (\downarrow) \\ &\lambda\ (\downarrow) \rightarrow \text{misclassified points } (\uparrow) \end{aligned}}$$

(d) θ^*, s^* $\rightarrow$ deleted.

sparsity is preserved.

we can use the sparsity pattern as our data (x), and we have our output (y). using these, we can reconstruct our θ^*.

+0

5) Given:-

sample -> 500

candidate martin -> 355

candidate Astroid Belt -> 145

a) MLE

As there are only 2 candidates,

p(person voting astroid-belt) = p

p(person voting martin) = 1-p

Now, we converted our problem into a

bernoulli.

MLE -> find parameters which maximises

our data.

Likelihood, $L(p) = {}^{500}C_{145} \, p^{145} (1-p)^{355}$

For maximum, $\frac{dL(p)}{dp} = 0$

$$\frac{\partial}{\partial p}\left(500C_{145}\, p^{145}\, (1-p)^{355}\right) = 0$$

$$\Rightarrow 500C_{145}\left[145p^{144}(1-p)^{355} - p^{145}(1-p)^{354}\,355\right] = 0$$

$$\Rightarrow 145p^{144}(1-p)^{355} - p^{145}(1-p)^{354}\,355 = 0$$

$$\Rightarrow 145\,\cancel{p^{144}}\,(1-p)^{\cancel{355}} = 355\,p^{\cancel{145}}\,\cancel{(1-p)^{354}}$$

$$\Rightarrow \frac{p}{1-p} = \frac{145}{355}$$

$$\Rightarrow 355p = 145(1-p)$$

$$\Rightarrow 355p + 145p = 145$$

$$p = \frac{145}{500} = 0.29$$

$$\boxed{\text{MLE, people will vote for martin} = 0.29}$$

+10

(b) we need to calculate,

$$Pr\left(\frac{1}{m}\left|\sum_{i=1}^{m} x_i - E[x]\right| \leq \varepsilon\right)$$

$$= 1 - Pr\left(\frac{1}{m}\left|\sum_{i=1}^{m} x_i - E[x]\right| \geq \varepsilon\right)$$

$$\left.\begin{aligned} \hat{x} - s\% &\leq E[x] \\ E[x] - \hat{x} &\geq s\% \end{aligned}\right\} (1)$$

$$E[x] \leq \hat{x} + s\%$$

$$-s \leq \hat{x} - E[x]$$

Apply modulus on both sides,

$$\left.\begin{aligned} |-s| &\leq |\hat{x} - E[x]| \\ |E[x] - \hat{x}] &\geq s\% \end{aligned}\right\} (2)$$

Adding eqation (1) & (2), we

get

$$Pr\left(\frac{1}{m}\left|\sum_{i=1}^{m} x_i - E[x]\right| \geq \varepsilon\right) \leq 2\exp(-2m\varepsilon^2)$$

Here, $m = 500$

$\varepsilon = 0.05$

Pr (5% error margin)

$$= 1 - 2e^{-2m\varepsilon^2}$$

$$= 1 - 2e^{-\cancel{2}\times \cancel{500}^{\,5} \times \frac{5}{\cancel{100}} \times \frac{5}{\cancel{100}_{\,\cancel{20}\ 4\ 2}}}$$

$$= 1 - 2e^{-2.5}$$

$$= 0.835$$

$$\boxed{\text{Confidence that I am within 5\% error margin} = 83.5\%}$$

+10

Bays theorm

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Cheat

Sheet :-

conditional

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Linear Reg :-

$$Y = m_0 + m_1 x_1 + m_2 x_2 \cdots \cdots$$

$x \rightarrow m \times n$

$\theta \rightarrow n \times 1$

$y \rightarrow m \times 1$