Midterm 1

MACHINE LEARNING, SUMMER2018 QUARTER

Duration: 1 hours 45 minutes

Name: DU ID:

1. **This is closed book/notes exams**
2. **Please write your name and DU ID before starting the exam.**
3. **Show all the step of your answer and justify you answer/steps**
4. **Please write clearly and upto the point**

Problem 1.(8 =1+1+1+1+1+1+2 points.)
1a. What is the main condition/assumption in naive bayes.(Please try to write no more than a line. Infact two words are enough)

1b. is logistic regression model a generative or a discriminative model?

1c. If you are building classifier for 4 classes, What is base accuracy you want to beat if each class is equally likely.(Please just write % accuracy).

1d. Write the Bayes formula for $P(A|B)$ in joint probability form $P(A, B)$.

1e. What is maximum value of AUROC = Area Under the Receiver Operating Characteristic curve or AUC.
1f. is bigger value of AUC better(True/False).

1g Solve following problem.
You might be interested in finding out a patients probability of having liver disease if they are an alcoholic. Being an alcoholic is the test (kind of like a litmus test) for liver disease.

- A could mean the event Patient has liver disease. Past data tells you that 10% of patients entering your clinic have liver disease. $P(A) = 0.10$.
- B could mean the litmus test that Patient is an alcoholic. Five percent of the clinics patients are alcoholics. $P(B) = 0.05$.
- You might also know that among those patients diagnosed with liver disease, 7% are alcoholics. This is your $B|A$: the probability that a patient is alcoholic, given that they have liver disease, is 7%.

Calculate $P(A|B)$ i.e if the patient is an alcoholic, what is their chances of having liver disease.

Problem 2.(7=2+2+3 points.) Multi variate gaussian in D dimension is given by

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\Big(-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\Big)$$

where $|\Sigma|$ is determinant of matrix Σ.
2a. From the above formula, write down the symbol used for mean vector and covariance matrix. Mention their size too.

2b. Sketch the contours(points in 2D having same value or density) of a 2D gaussian with mean $\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and covariance matrix $\begin{bmatrix} 1 & 0 \\ 0 & .5 \end{bmatrix}$.

2c. In gaussian dicriminative analysis class condition densities are modeled as $p(\boldsymbol{x}|y = c) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu_C}, \Sigma_C)$. If Σ_C is diagonal

$$\begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_D^2 \end{bmatrix}$$

, this leads to naive bayes.
Assume $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix}$ and $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_D \end{bmatrix}$,
(If you are enrolled in COMP 3432). Simplify $(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma_c^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$. Keep $x_i - \mu_i$ type of terms together.
(If enrolled in COMP 4432) For $D = 2$ and 2×2 diagonal matrixΣ_c, show that 2 dimensional Gaussain can be factored into product of 2, one diemnstional Gaussian. Use the formula at beginning for Gaussian. Show your work.

Problem 3.(10 points.) Let scalar random variable x has probability density function $p(x|\sigma) = \frac{1}{2\sigma} \exp(-\frac{|x|}{\sigma})$. If we have N, I.I.D samples $\mathcal{D} = \{(x_i)\}_{i=1}^{i=N}$ of x, then compute the MLE estimate of σ.

Problem 4.(5 = 2 + 2 + 1 points.)
If we have lots of features (that is more x_is) and we want to perform feature selection in a linear regression problem, which kind of regularization method do we want to use and why? What about λ(bigger or smaller)) (Hint: ℓ_1 or ℓ_2?)

Problem 5.(10 points.) In logistic regression we model $P(y = 1|\boldsymbol{x}) = \sigma(\boldsymbol{w}^T \boldsymbol{x}) = \frac{1}{(1+\exp(-\boldsymbol{w}^T \boldsymbol{x}))}$ where $\boldsymbol{x}, \boldsymbol{w} \in R^d$. Prove that decision boundary in a logistic regression is a line in D dimension. (Hint. Equate class 1 and 0 probability and simplify. Only linear (not square, cubic or higher degree terms) term should be in the final equation.)