Midterm practice

SUMMAER 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**

Problem 1.(.5 points each.)
1a. What is the difference in supervised and unsupervised machine learning.

1b. Why are generative model called generative and discriminative model discriminative?

1c. Given some observation $\mathcal{D}$ write the M.L.E formualtion of estimation of parameters θ and MAP estimation of parameters θ.

1d. What is the set of values poisson random variable takes(called support).

1e. Does strictly convex function has unique global minumum.(yes/no).

1f. Conditional independence means

$$P(X, Y|Z) =$$

1g. In linear regression, which norm does feature selection(ℓ_1 *or* ℓ_2)

Problem 2.(2+2+1+.5 points.)

2a. Let $\boldsymbol{x} \in \{1, \cdots, K\}^D$, i.e $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$ and $x_i \in \{1, \cdots, K\}$. In generative model we need to specify class conditional distribution $P(\boldsymbol{x}|y = c)$. If we don't assume conditional independence on features, given class label how many parameters we need to estimate.

2b. If we assume conditional independence on features given class label, how may parameters we need to estimate.

2c. Assuming conditional independence on feature given class label leads to Naive Bayes classifier. Write right hand side of following equation for naive bayes classifier.
$p(\boldsymbol{x}|y = c, \boldsymbol{\theta}) =$
2d. If you have less data, which model(model in 2a or 2b(naive Bayes)) is likely to give you less test set error. Explain in no more than 1(preferred) or 2 line.

Problem 3.(4= (2+2) points.) Let scalar x be drawn from $\mathcal{N}(\mu_i, \sigma^2) = \frac{1}{\sqrt{(2\pi)}\sigma} \exp(\frac{(x-\mu)^2}{-2\sigma^2})$ (1-d Gaussian distribution). If we have N, I.I.D samples $\mathcal{D} = \{(x_i)\}_{i=1}^{i=N}$, then compute the MLE estimate of μ and σ.

Problem 4(2 points) In linear regression $y = \boldsymbol{w}^T \boldsymbol{x} + \epsilon$ estimate of $\boldsymbol{w}$ is $\hat{\boldsymbol{w}} = (X^T X)^{-1} X^T y)$. Hence residual vector e against fitted line is $\boldsymbol{e} = y - X\hat{\boldsymbol{w}}$. Show that residual vector is orthogonal to columns of X. Note X contains observation $\boldsymbol{x_i}$ along rows.