ComS 573: Machine Learning
Spring 2020

Take-home Final Exam
Due Wednesday, May 6, 2020, 11:59 pm on Canvas

Instructions

- Please give all the necessary steps in your answers but be concise and precise.
- Please consult the instructor if you have difficulty *understanding* any of the problems.
- You must work independently, *no discussion or collaboration with others allowed*.

Problem	Worth	Score
1	20	
2	10	
3	10	
4	10	
5	5	
6	5	
7	13	
8	17	
9	10	
Total	100	

1. (20 pts.) Consider a two category classification problem with one dimensional discrete feature x. Assume that the priors are $P(\omega_1) = 1/2$ and $P(\omega_2) = 1/2$. Let the domain of x be nonnegative integers $\{0, 1, 2, \ldots\}$. Let the class-conditional probabilities have the form

$$p(x|\omega_i) = e^{-\lambda_i} \frac{\lambda_i^x}{x!}, \quad x = 0, 1, 2, \ldots$$

where λ_i for $i = 1, 2$ are unknown parameters.

   (a) (10 pts.) Suppose that samples are drawn independently according to $p(x|\omega_1)$ and $p(x|\omega_2)$ and we get dataset $D_1 = \{1, 5\}$ and $D_2 = \{3, 9\}$ for ω_1 and ω_2, respectively. Derive the maximum-likelihood estimate for λ_1 and λ_2.

   (b) (10 pts.) Let λ_{ij} be the loss incurred for deciding ω_i when the true category is ω_j. Assume $\lambda_{11} = 0$, $\lambda_{12} = 1$, $\lambda_{21} = 3$, $\lambda_{22} = 0$. Derive the Bayes decision rule for the minimum risk classification.

2. (10 pts.) Assume we are given 300 labeled examples to train a decision tree classifier and report its performance. Describe how to use 3-fold cross-validation to estimate the classification accuracy of the learned classifier.

3. (10 pts.) Consider using a neural network for a binary classification problem such that given an input $\vec{x}_k$ the single output of the network $y(\vec{x}_k; W)$ corresponds to the posterior probability $P(\omega_1|\vec{x}_k)$. Assume we have i.i.d data $\{(\vec{x}_k, t_k)\}_{k=1}^n$, where $t_k \in \{0, 1\}$, $t_k = 1$ if $\vec{x}_k$ is labeled to category ω_1, and $t_k = 0$ if $\vec{x}_k$ is labeled to category ω_2, but there is a probability ϵ that the class label on a training data point has been incorrectly set. Derive the error function corresponding to the negative log (conditional) likelihood $p(t_1, \ldots, t_n|\vec{x}_1, \ldots, \vec{x}_n)$.

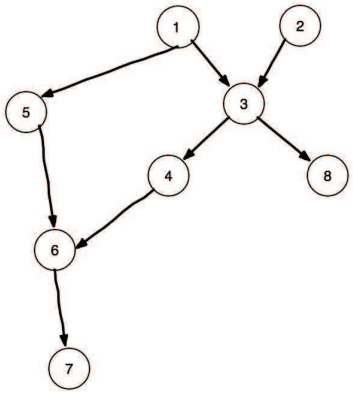
4. (10 pts.) We want to train a neural network to solve a classification problem. Discuss two possible approaches for controlling the overfitting issue.

5. (5 pts.) Given valid kernels $K_1(\vec{x}, \vec{x}')$ and $K_2(\vec{x}, \vec{x}')$, prove that the following will also be a valid kernel:

$$K(\vec{x}, \vec{x}') = K_1(\vec{x}, \vec{x}') + K_2(\vec{x}, \vec{x}')$$

6. (5 pts.) Both Bagging and Boosting work by collecting a set of base classifiers and using a (weighted) majority voting to classify new examples. What are the key differences between Bagging and Boosting methods?

7. (13 pts.) Consider the following Bayesian network



   (a) (4 pts.) List the local Markovian assumptions asserted by the network structure (that is, the assumptions under which the graph is a Bayesian network).

   (b) (6 pts.) True or false? Please explain if false.

      i. dsep(2,4,5)

      ii. dsep(2,{1,8},5)

      iii. dsep(2,{3, 7},{4,8})

   (c) (3 pts.) Which edge directions could be reversed in DAGs that are independence (d-separation) equivalent to the above DAG?

8. (17 pts.) Jack has two coins C_1 and C_2 with p_1 and p_2 as their corresponding probabilities of landing heads. First Jack randomly picks a coin (50-50 chance) and flips it. From then on he decides which coin to flip next as follows: 60% chance he will flip the same coin, with 40% chance of flipping the other coin. Jack performed the flip three times with given outcomes, e.g. (tail, head, tail). We want to determine which three coins were most likely flipped in turn (e.g., (C_1, C_1, C_2)).

   (a) (15 pts.) Describe a Bayesian network to solve this problem (i.e., to help determining which three coins were most likely flipped in turn given the outcomes of the three flips). Please specify the variables, their domains, the structure of the model, model parameters, and so on.

   (b) (2 pts.) What is the probabilistic query to ask the model (to determine most likely which three coins were flipped in turn given the three outcomes)?

9. (10 pts.) K-means is an iterative algorithm that initializes the cluster centers randomly. Is K-means algorithm guaranteed to converge regardless of how the cluster centers are initialized? Why?