Department of Computer Science & Engineering
Indian Institute of Technology Kharagpur
Mid-semester Examination, Autumn 2015
CS60050: Machine Learning

Full Marks: 100 **Time: 2 Hrs**

Answer ALL questions. You can use calculators.

1. Which of the following is true or false. Explain with reasons (max 2-3 sentences): **3+3+3=9**

   (a) Root mean square error is a good performance measure for multi-class classification problem.

   (b) Cross-validation is expected to reduce the variance in the estimate of error rate of a classifier.

   (c) Training set error will initially decrease and then increase as we increase the parameter C in soft-margin SVM training.

2. Write the soft-margin SVM problem. Write the KKT conditions for this problem. Derive the dual for soft-margin SVM. **10**

3. Consider the following *balloons* dataset: **10**

Color	Size	Act	Age	Inflated
YELLOW	SMALL	STRETCH	ADULT	T
YELLOW	SMALL	STRETCH	CHILD	T
YELLOW	SMALL	DIP	ADULT	T
YELLOW	SMALL	DIP	CHILD	T
YELLOW	LARGE	STRETCH	ADULT	T
YELLOW	LARGE	STRETCH	CHILD	F
YELLOW	LARGE	DIP	ADULT	F
YELLOW	LARGE	DIP	CHILD	F
PURPLE	SMALL	STRETCH	ADULT	T
PURPLE	SMALL	STRETCH	CHILD	F
PURPLE	SMALL	DIP	ADULT	F
PURPLE	SMALL	DIP	CHILD	F
PURPLE	LARGE	STRETCH	ADULT	T
PURPLE	LARGE	STRETCH	CHILD	F
PURPLE	LARGE	DIP	ADULT	F
PURPLE	LARGE	DIP	CHILD	F

Compute the parameters of Naive Bayes classifier for predicting *inflated*; and the training set error.

4. Define shattering and VC dimension. What is the VC dimension of the function class: circles in R^2 whose interiors are class 1 and exteriors are class 2. Give crisp reasons. There will be marks for conciseness. **15**

5. In the context of logistic regression, define the prediction and loss functions. Show that if the class conditional densities, $p(x|c_k)k = \{1, 2\}$, are gaussian, with equal co-variance, the posterior distribution $p(c_k|x)$ is a logistic. **10**

6. Consider the following dataset: **15**

   $\{(-, (1, -4)), (+1, (5, -1)), (+, (3, 3)), (+, (-1, 5)), (-, (-5, 2)), (-, (-2, -2))\}$

   Run 5 iterations of IRLS for finding the parameter w for fitting a logistic regression model without the bias term, w_0.

7. Define the expected squared loss function for regression problem $y = f(x), y \in R, x \in R^d$. Derive the bias- variance decomposition of the expected squared loss function, from first principles. **7+10+10=27**

   (a) Consider the regression problem where $x \in [0, 1]$, and $y = x^2$. Generate the dataset, D, taking 11 equidistant points at distance 0.1 starting with $x = 0$, and computing $y = x^2$. Learn the predictor function $y(x) = wx$ from the above dataset.

   (b) Calculate the bias considering $E_D[y(x)] = w'x$ and expected squared error with $y(x) = wx$, considering uniform distribution for x in the range $[0, 1]$. Which component in the bias-variance decomposition is the difference and why ?

■