수리확률론특강 2017년도 2학기 중간고사

2017학년 2학기

Date: 2017/10/19

Professor: 최형인 학번: 이름:

1. (20 points) Let $\varphi(x, \mathfrak{D})$ be a predictor in which the dataset $\mathfrak{D}$ is regarded as a random variable.
   (a) In case $\varphi(x, \mathfrak{D})$ is a regressor, give the definition of the aggregate regressor $\varphi_a(x)$.
   (b) In case $\varphi(x, \mathfrak{D})$ is a classifier, give the definition of the aggregate classifier $\varphi_a(x)$.
   (c) In case $\varphi(x, \mathfrak{D})$ is a regressor, show that
   $$E|\varphi_a(x) - f(x)|^2 \leq E_{\mathfrak{D}} E|\varphi(x, \mathfrak{D}) - f(x)|^2,$$
   where $f(x)$ is the optimal regression (true) function.
   (d) Using (c), explain why the aggregation reduces overall generalization error.

2. (10 points)
   (a) Give the definition of Bayes classifier.
   (b) Show that the Bayes classifier has the smallest generalization error among all classifiers.

3. (20 points) Suppose we toss a coin n times and the head comes up n_1 times.
   (a) What is the maximum likelihood estimator of the probability of the head's coming up?
   (Just give the answer. You don't need to give any justification.)
   (b) Suppose we use the Bayesian parameter estimation for the Bernoulli distribution with the beta distribution $\beta_{a,b}$ as its conjugate prior.
   What is the MAP estimator of the head's coming up?
   (c) What happens if $a = b$ and $a, b \to \infty$.
   (d) Show that for fixed a, b, if $n \to \infty$, then the MAP estimator converges to the maximum likelihood estimator.

4. (10 points) The error of ridge regression is
   $$E = \frac{1}{2}|X\theta - Y|^2 + \frac{1}{2}\lambda|\theta|^2$$
   (a) Find the normal equation.
   (b) Show this normal equation always has a unique solution if $\lambda > 0$.

5. (15 points)

   (a) Explain what is validation and what it is for.

   (b) Explain what is k-fold cross validation and how one uses it in model selection.

6. (10 points) Suppose we are given the following confusion matrix.

		Prediction	
		P	N
Actual	P	350	50
	N	100	500

   Compute the following.

   (a) TPR, FPR, TNR, FNR

   (b) Precision, Recall

   (c) F_1

7. (15 points) Suppose the maximum likelihood estimator is $\hat{\theta} = (\hat{w}, \hat{b})$. Let $P(y = k|x) = \text{softmax}_k(H_1(x), \cdots, H_K(x))$ where $H_j(x) = H_j(x_1, \cdots, x_d) = \hat{w_j} \cdot x + \hat{b_j} = \hat{w_{j1}} x_1 + \cdots + \hat{w_{jd}} x_d + \hat{b_j}$, for $j = 1, \cdots, K$.
   If $d = 2$ and $K = 4$, and H_1, H_2, H_3 are configured as in the following figure. Draw the decision region for the output $y \in \{1, 2, 3, 4\}$.