ELL 409

Minor Question Paper

28/09/2022

1

Consider a linear model of the form $y(x, w) = w_0 + \sum_{i=1}^{D} w_i x_i$, together with a sum-of-squares error function of the form $E_D(w) = (1/2) * \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2$. Now suppose that Gaussian noise ϵ_i with zero mean and variance σ^2 is added independently to each of the input variables x_i. By making use of E[i]=0 and E[ij] $= \delta_{ij}\sigma^2$, show that minimizing E_D averaged over the noise distribution is equivalent to minimizing the sum-of-squares error for noise-free input variables with the addition of a weight-decay regularization term, in which the bias parameter w_0 is omitted from the regularizer.

5 Marks

2

The dataset of pass/fail in an exam for 5 students is given in the table below. If we use Logistic Regression as the classifier and assume the model suggested by the optimizer will become the following for Odds of passing a course: $log(Odds) = -64 + 2 \times hours$

$Studyhours$	$Result(Pass = 1, Fail = 0)$
29	0
15	0
33	1
28	1
39	1

1. Calculate the probability of pass for students who studied for 33 hours?
2. How many hours should students at least study so that the probability of their passing the course is more than 95%?

5 Marks

3

For the following functions K(x,z), state if it is a kernel or not. If the function is a kernel, then write down its feature map. If it is not a kernel, prove that it is not one.

1. $x = [x_1, x_2]$, $z = [z_1, z_2]$, x_1, x_2, z_1, z_2 are real numbers. $K(x, z) = x_1 z_2$.
2. Let $x=[x_1, \ldots, x_d]$, $z=[z_1, \ldots, z_d]$, x_is and z_is are real numbers. $K(x,z)=1\langle x,z\rangle$.
3. $x = [x_1, \ldots, x_d]$, $z = [z_1, \ldots, z_d]$, and f is a function. $K(x,z) = f(x_1, x_2)f(z_1, z_2)$.

4. x=[x_1,...,x_d], z[z_1,...,z_d], x_is and z_is are integers between 0 and 100. $K(x,z)=\sum_{i=1}^{d}\min(x_i,z_i)$.
5. x = [x_1,...,x_d], z = [z_1,...,z_d], x_is and z_is are real numbers.
   $K(x, z) = (1 + x_1z_1)(1 + x_2z_2)\dots(1 + x_dz_d)$
6. x=[x_1,...,x_d],z=[z_1,...,z_d],x_is and z_is are integers between 0and 100. $K(x,z)=\sum_{i=1}^{d}\max(x_i,z_i)$.
7. x are z are documents with words from some dictionary D. K(x,z) is the number of words that occur in both x and z, where each unique common word is counted once.

5 Marks

4 Short Answer Type

Chooose the correct answer and give explanation to the one chosen *5 Marks*

4.1

An experimental analysis was performed for bias variance estimation on a regularized version linear regression model. Which of the conclusions thus deduced are true and why?

(a) Upon increasing the regularization parameter (λ), the model starts underfitting the data, whilst variance goes down.

(b) Upon decreasing the regularization parameter (λ), the model starts overfitting the data whilst bias goes up.

(c) Upon increasing the regularization parameter (λ), the model starts underfitting the data, whilst bias goes down.

(d) Upon decreasing the regularisation parameter (λ), the model starts overfitting the data whilst variance goes up.

4.2

Consider the following possible choices of error function in training a linear regressor model for preforming regression: (I) cross-entropy error, (II) Sum-of-squares error (III) Absolute error. Which of the following is/are true and why:

(a) (I) may not be so useful as the problem at hand is a regression problem.

(b) Both (II) and (III) will give the same result.

(c) (III) is problematic because it is not differentiable.

(d) (II) is preferred.

(e) Any one can be used for gradient descent.

4.3

Consider the binary classification problem label (y) satisfies, y∈{0, 1} and prior probability P(y=0) = π Now, if we model P(x|y = 1) to be the following distributions, which one(s) will cause the posterior P(y = 1|x) to have a logistic function form and why?

1. Gaussian
2. Poisson
3. Uniform
4. None of the above