Introduction to machine learning 89-511, Fall 2022 Home Assignment 2

Gal Chechik

Submission date: 2022-11-22

Submit Instructions

Submit to the 'submit' system one file:
- 'solutions.pdf' with your solutions to questions no. 1-5

1. (20 pts) Gaussian noise as a regularizer

For a set of samples $(x_1, y_1) \ldots, (x_n, y_n)$ with $x_i \in R$ and $y_i \in R$, we defined in class the problem of polynomial curve fitting as an error-minimization problem. Assume that the labels $\{y_i\}_{i=1}^N$ are i.i.d. and are drawn from a Gaussian distribution with a mean that is the prediction of the model $h_{\mathbf{w}}(x)$, and a standard deviation σ. Formally, $y_i \sim \mathcal{N}(h_{\mathbf{w}}(x_i), \sigma)$.

Show that maximizing the log of the likelihood $P(y_1, ...y_n|\mathbf{w})$ is equivalent to minimizing the squared error $Err(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^n (h_{\mathbf{w}}(x_i) - y_i)^2$ used in class. Recall that the likelihood of a dataset $S_n = \{x_1, \ldots, x_n\}$ is defined as $\mathcal{L} = P(S|\mathbf{w})$.
Hint: No need to compute any derivatives.

2. (12 pts) Bayesian decision theory

Show that for the squared cost function $\lambda(\hat{\theta}|\theta) = (\hat{\theta} - \theta)^2$, the optimal Bayes estimator is the conditional expectation.

$$\hat{\theta}_{SE} = E\left(\theta|S_n\right) = \int_{-\infty}^{\infty} \theta P(\theta|S_n) d\theta \quad .$$

Hint: The derivative operator is a linear operator, so it can switch order with the integral (you can use the Leibniz integral rule).

3. (20 pts) The effect of priors

Given a set of samples: $S_n = \{2, 2, 3, 4, 5\}$ drawn from a Poisson distribution. Find the conditional expectation

(a) Assuming a uniform prior.

(b) Assuming an exponential prior with parameter value 1 ($P(\theta) = \exp(-\theta)$). Compare to (a).

4. (28 pts) Maximum likelihood estimation

Let $S_n = \{x_1, \ldots, x_n\}$ be a random sample from the distributions below. Find the maximum likelihood estimate of the parameter θ.

(a) Poisson: $P(k|\theta) = \frac{\theta^k}{k!}\exp(-k)$ for $k = 0, 1, \ldots$.

(b) Exponential: $P(x|\theta) = \frac{1}{\theta}\exp(-x/\theta)$ for $x \geq 0$.

(c) $P(x|\theta) = \frac{1}{2}\exp(-|x-\theta|)$.

(d) $P(x|\theta) = \theta x^{\theta-1}$.

5. (20 pts) Bayesian decision boundary

Let X be an R.V. with a Gaussian distribution, where $P(x|\omega = 1) = \mathcal{N}(\mu_1, \sigma_1^2)$ and $P(x|\omega = 2) = \mathcal{N}(\mu_2, \sigma_2^2)$. σ_1 and σ_2 may not be identical. Assuming the prior is uniform $P(\omega = 1) = P(\omega = 2)$, find the decision boundary as a function of x, and the parameters μ_1, μ_2, σ_1, σ_2 .