

SML

February 2020

Assignment 3

Q1. Let \mathbf{x} be a d -dimensional binary vector with a multivariate Bernoulli distribution

$$P(\mathbf{x}|\theta) = \prod_{i=1}^d \theta_i^{x_i} (1 - \theta_i)^{1-x_i} \quad (1)$$

where $\theta = (\theta_1, \dots, \theta_d)^t$ is an unknown parameter vector, θ_i being the probability that $x_i = 1$. Show that the maximum likelihood estimate for θ is

$$\hat{\theta} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k \quad (2)$$

Q2. Show that if our model is poor, the maximum likelihood classifier we derive is not the best by exploring the following example. Suppose we have two equally probable categories. Further, we know that $p(x|\omega_1)\tilde{N}(0, 1)$ but assume that $p(x|\omega_2)\tilde{N}(\mu, 1)$. Imagine however that the true underlying distribution is $p(x|\omega_2)\tilde{N}(1, 10^6)$.

a. What is the value of maximum likelihood estimation μ_{ML} in our poor model, given a large amount of data?

b. What is the decision boundary arising from this maximum likelihood estimate in the poor model?

c. Give an expression for the optimal Bayes decision boundary given the true underlying distributions $p(x|\omega_1)\tilde{N}(0, 1)$ and $p(x|\omega_2)\tilde{N}(\mu, 1)$. Compare this with part b.

Q3. Suppose we employ a novel method for estimating the mean of a data set $\mathcal{D} = x_1, x_2, \dots, x_n$; we assign the mean to be the value of the first point in the set ie. x_1 .

a. Show that this method is unbiased.

b. State why this method is nevertheless highly undesirable.

Ungraded

Q4. Derive MLE for Binomial distribution $X \sim \text{Bin}(N, \mu)$

$$P(X = m) = (N_{Cm}) \mu^m (1 - \mu)^{N-m} \quad (3)$$

The distribution gives the probability of observing m successes (say heads) in N independent Bernoulli trials. Since the heads can appear anywhere, N_{Cm} is the normalizing factor.