CS 760: Machine Learning Spring 2024

Homework 6: Frequentists vs Bayesians

AUTHORS: Jed Pulley

DO NOT POLLUTE! AVOID PRINTING, OR PRINT 2-SIDED MULTIPAGE.

Problem 6.1. Frequentist (MLE)

To find the MLE of $p*$, we first start with the likelihood function:

$$\mathbb{P}(p*) = \prod_{i=1}^{n} (p^{\mathsf{x_i}}(1-p)^{1-\mathsf{x_i}})$$

Then we take the log of the likelihood function:

$$log(\mathbb{P}(p^*)) = logp \sum_{i=1}^{n} \mathsf{x_i} + log(1-p) \sum_{i=1}^{n} (1-\mathsf{x_i})$$

Using our optimization 101 technique, we get the derivative and set it equal to zero:

$$p_{MLE} = \frac{1}{n} \sum_{i=1}^{n} \mathsf{x_i}$$

Which we recognize to just be the mean.

Problem 6.2. Bayesian (MAP)

We use our MAP formula $p_{MAP} = \arg\max_p \mathbb{P}(X|p)$ as our starting point. Using Bayes Rule, we can rearrange it as such:

$$p_{MAP} = \arg\max_p \frac{\mathbb{P}(X|p)\mathbb{P}(p)}{\mathbb{P}(X)}$$

Since we're maximizing for p, we can ignore the $\mathbb{P}(\mathbb{X})$ term as it doesn't depend on p:

$$p_{MAP} = \arg\max_p \mathbb{P}(X|p)\mathbb{P}(p)$$

This is very similar to our MLE statement above, with the exception of our prior term $\mathbb{P}(p)$ which we're assuming is information gathered about a previous event. Notably, if there is no prior information, our MAP estimate is equal to our MLE.

Given our prior $\mathbb{P}(p)$ being modeled as $Beta(\alpha, \beta)$, we can show our prior below:

$$\mathbb{P}(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

Using the prior information, the likelihood of our data (we'll say $X = [\mathsf{x}_1, \mathsf{x}_2, ..., \mathsf{x}_\mathsf{N}]^\mathsf{T}$) can be written as:

$$\mathbb{P}(X|p) = p^{\sum_{i=1}^{N} \mathsf{x}_i}(1-p)^{N-\sum_{i=1}^{N} \mathsf{x}_i} = p^{\mathsf{1}^\mathsf{T}\mathsf{X}}(1-p)^{N-\mathsf{1}^\mathsf{T}\mathsf{X}}$$

Simplifying and matching with the form of a Beta distribution, we find our MAP estimator to be:

$$\hat{p}_{MAP} = \frac{\sum_{i=1}^{N} \mathsf{x}_\mathsf{i} + \alpha - 1}{N + \alpha + \beta - 2}$$

Problem 6.3

(a)

(b)

Problem 6.4

(a)

(b)

Problem 6.5