CS 475 Machine Learning: Homework 1 Analytical (50 points)

Assigned: Friday, September 6, 2024
Due: Friday, September 20, 2024, 11:59 pm US/Eastern

Trevor Black (tblack20)

Instructions

We have provided this LaTeX document for turning in this homework. We give you one or more boxes to answer each question. The question to answer for each box will be noted in the title of the box. You can change the size of the box if you need more space.

Other than your name, do not type anything outside the boxes. Leave the rest of the document unchanged.

Do not change any formatting in this document, or we may be unable to grade your work. This includes, but is not limited to, the height of textboxes, font sizes, and the spacing of text and tables. Additionally, do not add text outside of the answer boxes. Entering your answers are the only changes allowed.

We strongly recommend you review your answers in the generated PDF to ensure they appear correct. We will grade what appears in the answer boxes in the submitted PDF, NOT the original latex file.

1 Maximum Likelihood [10 pts]

1. (5 pts) Given a dataset of n independently and identically distributed data points $\{x_i\}_{i=1}^n$, where each x_i is drawn from a Gaussian distribution, $x_i \sim \mathcal{N}(\mu, \sigma^2)$. Show that the MLE estimator of the mean is:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

Likelihood

$\mathcal{L}(\mu, \sigma^2 | \{x_i\}_{i=1}^n) = \prod_{i=1}^n p(x_i|\mu, \sigma^2)$

$p(x_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x_i-\mu)^2}{2\sigma^2})$

$\mathcal{L}(\mu, \sigma^2 | \{x_i\}_{i=1}^n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x_i-\mu)^2}{2\sigma^2}) = (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp(-\Sigma_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2})$

Log likelihood

$\log \mathcal{L}(\mu, \sigma^2 | \{x_i\}_{i=1}^n) = \ell(\mu, \sigma^2 | \{x_i\}_{i=1}^n) = \log((\frac{1}{\sqrt{(2\pi\sigma^2)}})^n \exp(-\Sigma_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}))$

$\ell(\mu, \sigma^2 | \{x_i\}_{i=1}^n) = n\log(\frac{1}{\sqrt{2\pi\sigma^2}}) - \frac{1}{2\sigma^2}\Sigma_{i=1}^n (x_i-\mu)^2 = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\Sigma_{i=1}^n (x_i-\mu)^2$

Maximize log likelihood

$\frac{d}{d\mu}\ell(\mu, \sigma^2 | \{x_i\}_{i=1}^n) = \frac{1}{\sigma^2}\Sigma_{i=1}^n (x_i-\mu) = 0$

$0 = \frac{1}{\sigma^2}\Sigma_{i=1}^n (x_i-\mu) = \Sigma_{i=1}^n (x_i-\mu) = -n\mu + \Sigma_{i=1}^n (x_i)$

$\mu = \frac{1}{n}\Sigma_{i=1}^n x_i$

2. (5 pts) Show that the MLE estimator of the variance is:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu})^2 \tag{2}$$

TODO

2 Conditional Independence [20 pts]

Let us create a probabilistic model where the presence/absence of each of these two symptoms, cough and fever, are conditionally independent given the presence/absence of pneumonia. Using this data for the empirical probabilities of our model, answer the following questions.

1. (5 pts) Find the probability that someone has both a cough and a fever.

2. (5 pts) Find the probability that someone has pneumonia given that they have a fever but no cough.

3. (10 pts) Given assumptions described above, how many parameters do we need to specify the joint distribution $p(fever, cough, pneumonia)$?

3 Bayesian Reasoning [20 pts]

1. (5 pts) Define what conjugate priors are, and explain why they are useful.

2. (10 pts) Show that the Gamma distribution is a conjugate prior of the exponential distribution. That is, show that if $x \sim \text{Exp}(\lambda)$ and $\lambda \sim \text{Gamma}(\alpha, \beta)$, then $p(\lambda|x) \sim \text{Gamma}(\alpha^*, \beta^*)$ for some α^*, β^*.

3. (5 pts) Derive the maximum a posteriori (MAP) of λ under the $Gamma\,(\alpha, \beta)$ prior.