CSE 5523 Homework 3: Parameter Estimation

Alan Ritter

Problem 1

Consider the following data set on lung diseases. Your goal is to build a Naïve Bayes classifier that predicts whether a person has Bronchitis or Tuberculosis, given his/her symptoms.

Disease	X-ray Shadow	Dyspnea	Lung Inflamation
Bronchitis	Yes	Yes	Yes
Bronchitis	Yes	Yes	Yes
Bronchitis	No	No	Yes
Tuberculosis	Yes	Yes	No
Tuberculosis	Yes	No	No
Tuberculosis	No	No	Yes

1. *(1 point)* List the distributions that would be learned if you use a maximum likelihood estimate (MLE) to estimate the parameters of a Naïve Bayes model from this data (e.g. $P(\text{Dyspnea}|\text{Bronchitis}) = ?$). Include all of the parameters. Show your work.

2. *(1 point)* Based on your learned model, diagnose a patient with the following symptoms (show your work):

X-ray Shadow	Dyspnea	Lung Inflamation
Yes	No	Yes

Problem 2

Assume you are given a dataset of n real numbers $D = \{X^1, X^2, \ldots, X^n\}$, $X^i \in \mathcal{R}$. Derive the maximum likelihood mean, μ and variance, σ parameters of a 1-dimensional Gaussian distribution.

1. *(1 point)* Write down the log-likelihood of D as a function of μ and σ, $\mathcal{L}(\mu, \sigma)$.

2. *(1 point)* Compute the partial derivative of $\mathcal{L}(\mu, \sigma)$ with respect to μ, equate to zero and solve for μ.

3. *(1 point)* Compute the partial derivative of $\mathcal{L}(\mu, \sigma)$ with respect to σ, equate to zero and solve for σ.

Problem 3

Consider a training sample of inputs $X^1, X^2, \ldots, X^n$ and outputs $Y^1, Y^2, \ldots, Y^n$, where inputs are real-valued vectors $X^i \in \mathcal{R}^V$ and outputs are binary $Y^i \in [0, 1]$.

Recall (from the slides presented in class) that the conditional log likelihood can be written as follows:

$$\begin{aligned}
\mathcal{L}(W) &= \sum_l Y^l \log P(Y=1|X^l, W) + (1 - Y^l) \log P(Y=0|X^l, W) \\
&= \sum_l Y^l \log \frac{P(Y=1|X^l, W)}{P(Y=0|X^l, W)} + \log P(Y=0|X^l, W) \\
&= \sum_l Y^l \left(w_0 + \sum_{i=1}^{V} w_i X_i^l \right) - \log \left(1 + \exp(w_0 + \sum_{i=1}^{V} w_i X_i^l) \right)
\end{aligned}$$

1. *(2 points)* Show that the partial derivative of $\mathcal{L}(W)$ with respect to w_i is as follows:

$$\frac{\partial \mathcal{L}(W)}{\partial w_i} = \sum_l X_i^l (Y^l - P(Y=1|X^l, W))$$

2. *(2 points)* Now, assume a zero-mean Gaussian prior over the weights:

$$P(w_i) = \mathcal{N}(0, \sigma)$$

Write down the expression for the posterior distribution over w_i, and derive the gradient (e.g. that can be used for estimating MAP parameters in gradient decent).