Name: Brandon Botzer

Date: 1/14/2024

Class: Penn State – AI 879

Homework 4: Classification and Regression

Problem 6.1

Consider the following problems.

i Determining the gender of an image of a face.

ii Determining the pose of the human body given an image of the body.

iii Determining which suit a playing card belongs to based on an image of that card.

For each case, try to describe the contents of the world state w and the data x. Is each discrete or continuous? If discrete, then how many values can it take? Which are regression problems and which are classification problems?

For 6.1.i:

The world state would be the gender of the image. With updated social understandings between the difference between sex and gender, this question is more likely asking to define the world state as the sex of an individual, male or female. Gender is culturally defined and based on societal expectations. In the end, the biological sexes, male and female, are biologically discrete world states. The data, x, is continuous and could be the shapes of facial features. This would bin this as a classification problem (this is further expanded upon in chapter 9).

For 6.1.ii:

The world state would be the position of head, neck, limbs, and torso of the body. This would be a continuous world state and data given that a limb can be set at any joint angle ($a \in \{0, 2\pi\}\ rad$). This would be a regression problem (it is also further expanded upon in chapter 8).

For 6.1.iii:

The world state, for a set of classical playing cards, would be the four discrete suits: hearts, diamonds, spades, clubs. For the data, x, there are 52 cards in a standard deck (assuming no jokers). The data is continuous as it could be the shape of the suit or use the color of a channel to narrow the suit down. Given the multiple classes (N = 4 world states), this would be a multi-class logistic regression problem (Prince, pg. 197), which is a classification problem.

Problem 6.2

Describe a classifier that relates univariate discrete data x ∈ {1 . . . K} to

a univariate discrete world state w ∈ {1 . . . M } for both discriminative and generative

model types.

Consider a model which identifies fruit (apples and oranges) (binary classification).

For the discriminative model:

We first define a probability distribution over the world state $w \in \{0,1\}$ (apple or orange) and make its parameters contingent on the data x. Given a discrete and binary world state, a Bernoulli distribution is used with the single parameter λ which would determine the probability of fruit type such that $P(w = 1) = \lambda$.

Since the data, x, is discrete, we will use a categorical distribution (such as a normalized histogram) where $P(x = k|\lambda_{\{1 \ldots K\}}) = \lambda_k$.

We could use the Maximum likelihood to compute λ_k.

You could also do this with a softmax function and a categorical distribution on the training data learned values (l) where:

$$\lambda = softmax(x) = \frac{e^{l}}{\sum_{j=1}^{K} e^{l_j}}$$

For the generative model:

To model $P(x|w)$, the form of $P(x)$ over the data is chosen and the distribution parameters are a function of the world state w. For this discrete binary classifier, we define a prior as a categorical distribution (Bernoulli) is used with the parameter vector λ being a function of the world state ($P(w) = Bern_w[\lambda_p]$) which gives $P(w = 1) = \lambda \; and \; P(w = 0) = 1 - P(w = 1)$. The prior parameter λ_p is learned through the training world states $\{w_i\}_{i=1}^{I}$.

Using Baye's rule, the posterior can be computed by:

$$P(w|x) = \frac{P(x|w)P(w)}{\sum_{w=0}^{1} P(w|x)P(w)} = \frac{P(x,w)}{P(x)}$$

Problem 6.3

Describe a regression model that relates univariate binary discrete data x ∈

{0, 1} to a univariate continuous world state w ∈ [−∞, ∞]. Use a generative formulation

in which P r(x|w) and P r(w) are modeled.

A risk formula (ie. credit score) could be a similar use case to this. A good or bad score could map to any given range of values $\{-\infty, \infty\}$ based on how the risk score is calculated.

The problem is structurally similar to the last but with a change in the denominator.

In total, the probability distribution over the data x could use the Bernoulli distribution where: $P(x|w, \theta) = Bern[\lambda]$.

A prior distribution would need to be chosen. One such could be a univariate normal with variance $\sigma^2$and mean μ: $P(w) = Norm_w[\sigma^2, \mu]$.

The parameters θ would need to be fit from training data examples for each class $\{x_i, w_i\}_{i=1}^{I}$. The parameter fitting can be done using a Sigmoid function, $\frac{1}{1+\, e^{\{\phi_0 + \phi_1 w)}}$

Using Baye's rule we have:

$$P(w|x) = \frac{P(x|w)P(w)}{\int_{-\infty}^{\infty} P(x|w)P(w)dw}$$