HW 3

Wednesday, October 27, 2021 7:03 PM

Problem 1:

$$E_{(x,y)}[l(f(x),y)] = \int_x \left\{ \int_y (z-y)^2 p(y|x) dy \right\} p(x) dx = 0$$

$$\frac{\partial}{\partial x}\left[\int_x \left\{ \int_y (z-y)^2 p(y|x) dy \right\} p(x) dx = 0 \right]$$

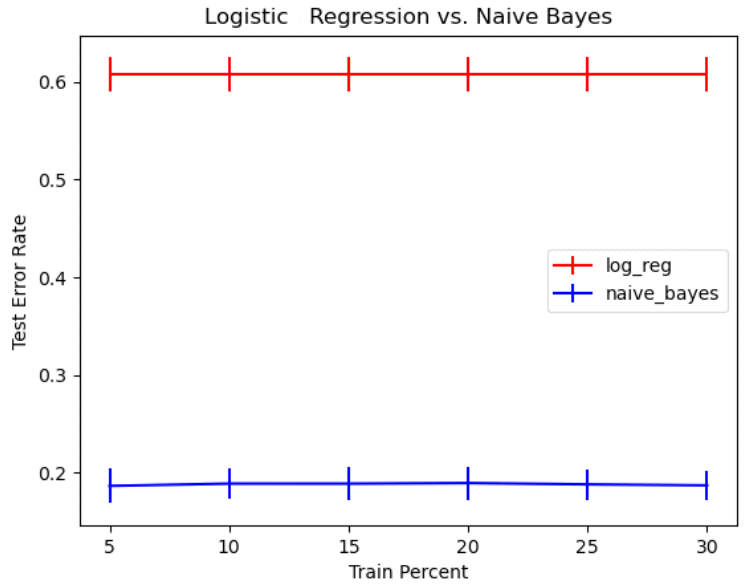
$$\int_x \left\{ \int_y \frac{\partial}{\partial x} (z-y)^2 p(y|x) dy \right\} p(x) dx = 0$$

$$\int_x \left\{ \int_y 2(z-y) p(y|x) dy \right\} p(x) dx = 0$$

$$2z = 2y\big(p(y|x)\big)$$

$$f(x) = E[y|x]$$

Problem 2



Approach for Logistic Regression:

Logistic regression using stochastic gradient descent. Formula found in lecture notes but specifically it is $\vec{\theta} \coloneqq \vec{\theta} - \alpha \nabla_{\vec{\theta}} J(\vec{\theta})$

Where $\nabla J_n(\vec{\theta}) = (f_{\vec{\theta}}(\vec{x}) - y_n)\overrightarrow{x_n}$ and $f_{\vec{\theta}}(\vec{x}) = \frac{1}{1+e^{-\vec{\theta}^T \vec{x}}}$

Theta was updated 1000 times and then used to predict the target value

Approach for Naive Bayes

Naive Bayes was formulated as such.

Using the 80-20 split the mean and standard deviation was calculated for each feature in each class.

Leaving us with $\overrightarrow{\mu_0}, \overrightarrow{\sigma_0}$ for $y = 0$ & $\overrightarrow{\mu_1}, \overrightarrow{\sigma_1}$ for $y = 1$

Then for each $\vec{x}$ in the testing split against the target value y I calculated

$P(\vec{x}|y = 1)P(y = 1)$ and

$P(\vec{x}|y = 0)P(y = 0)$

Where $P(\vec{x}|y = 0) \sim N(\vec{x}, \overrightarrow{\mu_0}, \overrightarrow{\sigma_0})$ and $P(\vec{x}|y = 1) \sim N(\vec{x}, \overrightarrow{\mu_1}, \overrightarrow{\sigma_1})$

Results:

Naive Bayes worked well. For each training percent it was roughly 80% accurate. With some Laplace smoothing to avoid divide by 0 errors the algorithm ran well and gave accurate results even using 5% of the training data.

Logistic regression did not perform as well. Which disagrees with the papers results. This is due to poor implementation of the algorithm. Things that could have changed this outcome would be training the data for much longer, and potentially using batch gradient descent to sacrifice speed for accuracy.

Remark to grader:

Logistic regression is implemented very poorly. I have followed the formula exactly as shown in the lecture notes following the binary classification but for some reason the gradient descent does not descend properly. I have no other explanation for my results other than that.