Matt Kinderman

Machine Learning: Assignment 4

Due on Wed, Feb 29, 2023 at 11:00am

Instructor: Dr. Anh Nguyen

Matt Kinderman

Problem 2

10 points Explain in 3-5 sentences why logistic regression is a *discriminative* classifier as opposed to Naive Bayes, which is a *generative* classifier. In other words, what makes logistic regression *discriminative* and what makes Naive Bayes *generative*?

Notes: Write down mathematics equations, if necessary, to support your answer.

Naive Bayes is considered a generative classifier because it models the joint probability distribution of the features and target variable to estimate P(X|Y) and P(Y). These probabilities use Bayes' theorem to calculate P(Y|X). On the other hand, logistic regression uses P(Y|X=x) to make predictions. Therefore, it is considered a discriminative model because it estimates the conditional probability of the target variable given the features.

Problem 3

10 points The prediction rule for a logistic-regression, binary classifier is if $P(y=1|\boldsymbol{x}) > P(y=0|\boldsymbol{x})$ then output 1 otherwise, output 0.

Assume that our $\boldsymbol{w} \in \mathbb{R}^2$ (i.e., there are two weights w_1 and w_2 only) and there is a bias $b \in \mathbb{R}$. The label $y \in \{0,1\}$. The input features $\boldsymbol{x} \in \mathbb{R}^2$.

Question: Derive the decision boundary equation for this classifier.

Hints: The decision boundary of this logistic regression classifier is a line at the point when

$$P(y=1|\boldsymbol{x}) = P(y=0|\boldsymbol{x})$$

From the above equation, continue deriving step-by-step (with justifications) to arrive at the equation for the linear decision boundary (of this logistic regression classifier) separating two classes ($y=0$ and $y=1$).

$$P(y=1|x) = \frac{1}{1+\exp\left(\sum w_i x_i + w_0\right)}$$

$$P(y=0|x) = \frac{\exp\left(\sum w_i x_i + w_0\right)}{1+\exp\left(\sum w_i x_i + w_0\right)}$$

$$\frac{P(y=0|x)}{P(y=1|x)} = \frac{\left(1+\exp\left(\sum w_i x_i + w_0\right)\right) * \exp\left(\sum w_i x_i + w_0\right)}{1+\exp\left(\sum w_i x_i + w_0\right)}$$

Simplify…

$$\frac{P(y=0|x)}{P(y=1|x)} = \exp\left(\sum w_i x_i + w_0\right)$$

Matt Kinderman

$$\ln\left(\frac{P(y=0|x)}{P(y=1|x)}\right) = \sum w_i x_i + w_0$$

$$\ln(1) = \sum w_i x_i + w_0$$

$$0 = w_0 + w_1 x_1 + w_2 x_2$$

Problem 4

10 points Given the classifier in Problem 2, derive the full conditional log likelihood function $L(\boldsymbol{w})$ that we want to maximize (i.e. MCLE not MAP). We have N examples (i.e., pairs of $(\boldsymbol{x}^i, y^i)$) in the training set.

$$L(\boldsymbol{w}) = \log_e \prod_{i=1}^{N} P(y^i|\boldsymbol{x}^i, \boldsymbol{w})$$

$$L(w) = \ln \sum_{i=1}^{N} P(\, y^i \,|x^i, w)$$

$$L(w) = \sum_{i=1} y^i \ln\, P(y^i = 1|x^i, w) + \left(1 - y^i\right) \ln P(y^i = 0|x^i, w)$$

$$L(w) = \sum_{i=1} y^i \ln \frac{P(y^i = 1|x^i, w)}{P(y^i = 0|x^i, w)} + \; P(y^i = 0|x^i, w)$$

$$L(w) = \sum_{i=1} y^i \left(w_0 + \sum_{j=1}^{n} w_j x_j^i \right) - \ln\left(1 + exp\left(w_0 + \sum_{j=1}^{n} w_j x_j^i \right)\right)$$

Matt Kinderman

Problem 5

10 points Extend the given classifier in Problem 2 to a single-label, 3-way classifier, i.e., given $\boldsymbol{x} \in \mathbb{R}^2$, predict one label for $y \in \{0, 1, 2\}$.
Hint: How many set of weights and biases are there in this case?

Question: Write down the full definition of the three below functions (i.e., in terms of exp and the parameters w_1, w_2,...so that someone can plug in the parameters and compute the output probabilities).

$$P(y=0|x) = \frac{\exp(w_0^T x)}{\exp(w_0^T x) + \exp(w_1^T x) + \exp\,(w_2^T x)}$$

$$P(y=1|x) = \frac{\exp(w_1^T x)}{\exp(w_0^T x) + \exp(w_1^T x) + \exp\,(w_2^T x)}$$

$$P(y=2|x) = \frac{\exp(w_2^T x)}{\exp(w_0^T x) + \exp(w_1^T x) + \exp\,(w_2^T x)}$$

Problem 6

10 points Given the classifier in Problem 2, extend the classifier to a single-label, 3-way **softmax** classifier. That is, given $\boldsymbol{x} \in \mathbb{R}^2$, predict one label for $y \in \{0, 1, 2\}$. Yet, use the **softmax** function instead of the **sigmoid**.

Question: Write down the full definition of the three below functions (i.e., in terms of exp and the parameters w_1, w_2,...so that someone can plug in the parameters and compute the output probabilities).

We will use the formula... $\sigma(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{k} e^{y_j}}$

$$y_i = w_0 + w_1^i x_1^i + w_2^i x_2$$

$$P(y=0|x) = \frac{e^{b_1 + w_1^1 x_1 + w_2^1 x_2}}{1 + e^{b_1 + w_1^1 x_1 + w_2^1} + e^{b_1 + w_1^2 x_1 + w_2^2}}$$

$$P(y=1|x) = \frac{e^{b_2 + w_1^2 x_1 + w_2^2 x_2}}{1 + e^{b_1 + w_1^1 x_1 + w_2^1} + e^{b_1 + w_1^2 x_1 + w_2^2}}$$

Matt Kinderman

$$P(y = 2|x) = \frac{1}{1 + e^{b_1 + w_1^1 x_1 + w_2^1} + e^{b_1 + w_1^2 x_1 + w_2^2}}$$