Likelihood Machine Learning Seminar II

Lecture 3: Logistic Regression

Likelihood Lab

XingYu Fu

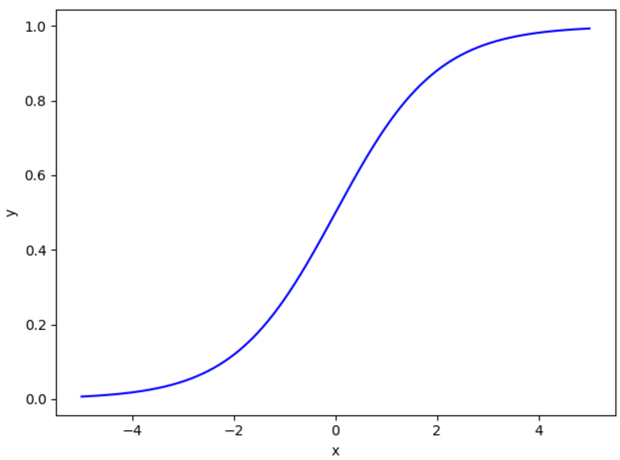
1. How Logistic Regression Predict?

Logistic Regression is a machine learning algorithm that is designed to solve classification problems with two classes. Formally, say we want to classify a feature vector $x \in R^n$ to two possible categories $\{0,1\}$. Logistic regression calculates the probability that x belongs to class 1 as:

$$\frac{1}{1+e^{-(w \cdot x+b)}}$$

, where $w \in R^n$ is the weight of logistic model and $b \in R$ is the bias term.

The function $f(z)=\dfrac{1}{1+e^{-z}}$ is called the logistic function and the graph of it looks like:



2. How to Train Logistic Regression?

The key to use logistic regression to conduct prediction is having good weight w and bias b that can combine the attributes of unknown feature vector x smartly. The training process of logistic regression is, in fact, a searching process of optimal model parameter w and b. The most popular algorithm to train a logistic model is Stochastic Gradient Descent (SGD), where we minimize the *loss* generated in each training sample respectively using gradient descent. Formally, say we have a training sample (x, y). The loss generated in this sample is defined as:

$$L(w, b) = (\frac{1}{1 + e^{-(w \cdot x + b)}} - y)^2$$

and we update w and b using gradient descent:

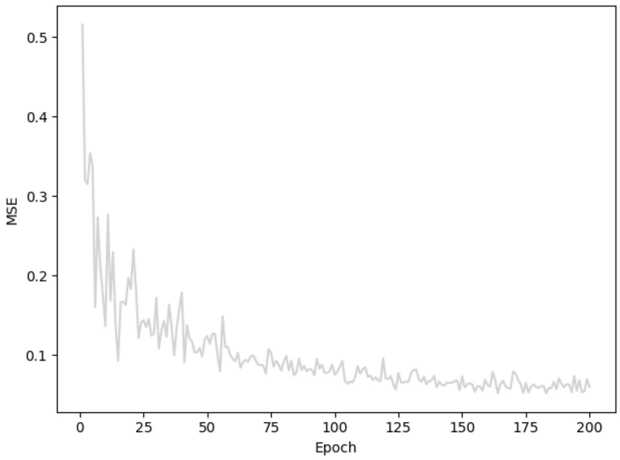
$$w = w - \eta \nabla L(w, b)$$

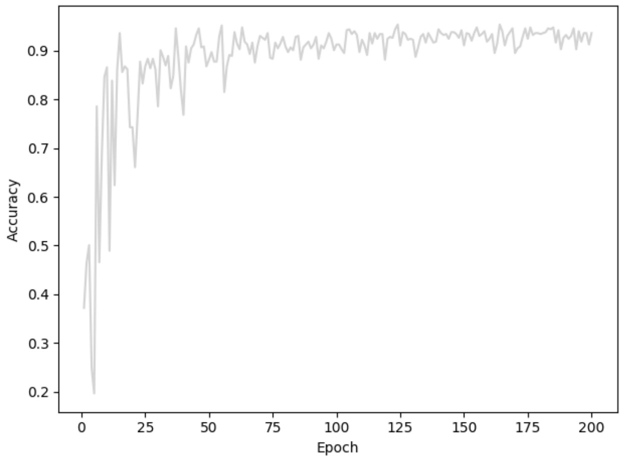
$$b = b - \eta \frac{\partial L(w, b)}{\partial b}$$

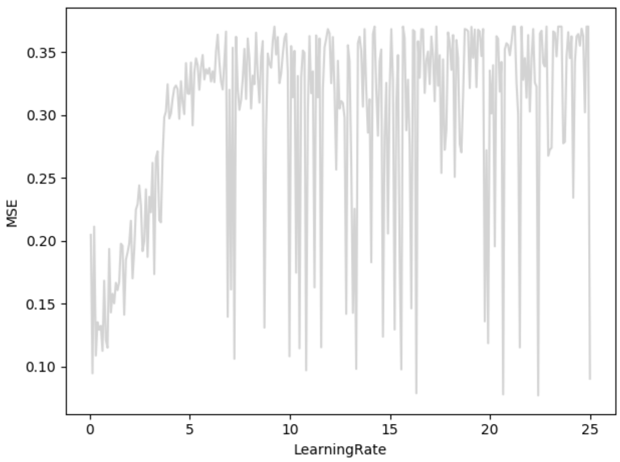
The constant $\eta \in (0,1)$ here is called the learning rate, controlling the speed of SGD update.

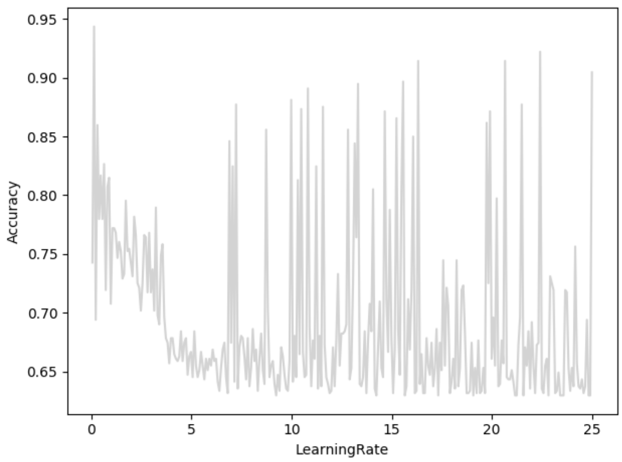
3. Example: Logistic Regression on Breast Cancer

We apply the logistic regression on the breast cancer Wisconsin (diagnostic) dataset. As we can see from the below figures, MSE and Accuracy of the logistic model improves as the training epoch increases (not always improve, look up the term: over-fitting), while we can also see that too large learning rate may exert a negative impact on the predictive performance.









4. Feature Importance

Understanding which feature contributes most in the predictive process is always an important work in practice since it can offer you intuition about how the system might work. In logistic model, the feature importance of the i^{th} attribute is defined as:

$$I(x_i) = |w_i|$$

In other words, larger absolute value of weight implies more important feature.