MACHINE LEARNING

Shoaib Farooq

Department of Computer Science

Instructor Information

Instructor	Shoaib Farooq	E-mail	m.shoaib1050@gmail.com
GitHub	https://github.com/SHOAIB1050		
YouTube	https://www.youtube.com/c/pakithub		
LinkedIn	https://www.linkedin.com/in/shoaib-farooq-b5b190105/		

Let’s Start .....

Lecture #7

GOALS

This Lecture Will Cover:

- **Classification Vs Logistic Regression**
- **Odds In Probability**
- **Logistic Regression**
- **Quantitative Evaluation of ML models.**

CLASSIFICATION VS LOGISTIC REGRESSION

Classification

- **Goal:** To categorize data into predefined classes or categories. For example, determining if an email is "spam" or "not spam."
- **Methods:** Various algorithms, including decision trees, support vector machines (SVM), k-nearest neighbors, and neural networks, fall under classification.
- **Output:** Usually a categorical label (like "yes" or "no") indicating the predicted class for a given input.
- **Probabilities:** Some classification models provide class probabilities (e.g., a probability that an email is spam), though not all do.

LOGISTIC REGRESSION

Logistic Regression

- **Goal:** Primarily used for binary classification tasks, logistic regression predicts the probability of an observation belonging to a specific class (e.g., "0" or "1").
- **Output:** Typically outputs a probability value (e.g., 0.85) for the positive class. This value can be converted to a binary class label using a threshold (e.g., if the probability is above 0.5, classify as "1"; otherwise, "0").
- **Limitations:** Logistic regression is linear, so it is best suited for linearly separable data and may struggle with more complex patterns without additional feature engineering or transformation.

LOGISTIC REGRESSION

- Linear regression models the relationship between a response variable and one or more explanatory variables.
- For categorical response variable with two possible values, Similar Regression Models can also be used
  - Spam or Not Spam
  - Patient Dies or Survives
  - Tumor Benign or Malignant.

LOGISTIC REGRESSION

- Classification, like regression, is a predictive task
  - But one in which the outcome takes only values across discrete categories;
- Classification problems are very common (more common than regression problems!)
- The objective function should be modified
- Fundamentals will be same as regression

LOGISTIC REGRESSION

- **Logistic Regression** (also called Logit Regression) is commonly used to estimate the probability for each class
  - What is the probability that this email is spam?
- Can a binary classifier be constructed using probability?
  - If the estimated probability > 50%, the instance belongs positive class
  - Otherwise, it belongs to the negative class

LOGISTIC REGRESSION

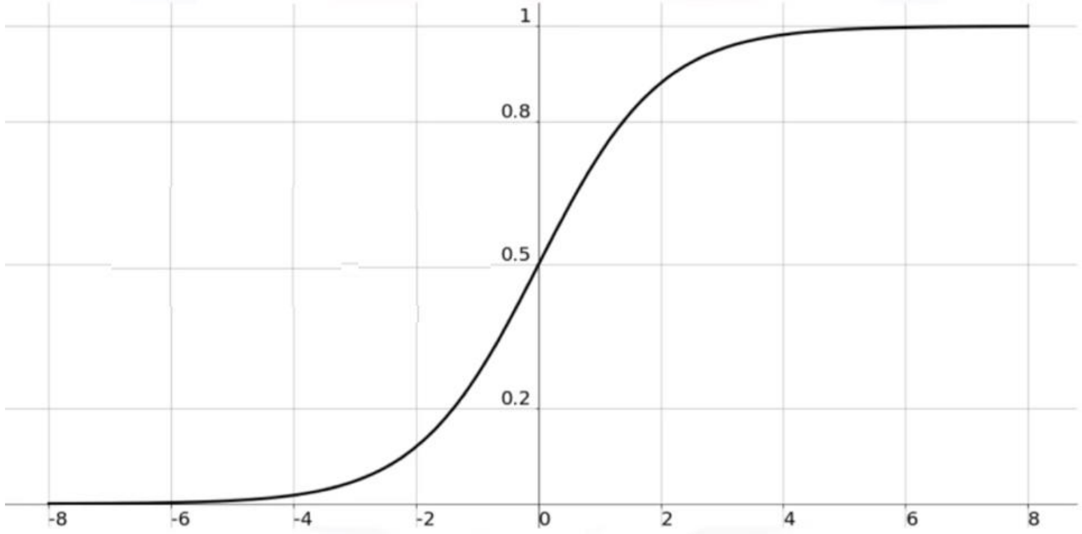
- It is a supervised method for classification
- “Logit” = “Log Odds”
- $p(y=0|x)$ or $p(y=1|x)$?

$$\log\left(\frac{p(x)}{1-p(x)}\right)$$

LOGISTIC REGRESSION

- Suppose $p(y=1|x) = p(x)$
- Sigmoid Function:

$$p(x) = \frac{1}{1+e^{-W^T x}}$$



LOGISTIC REGRESSION

- Linear Regression model, a Logistic Regression model computes a weighted sum of the input features (plus a bias term),

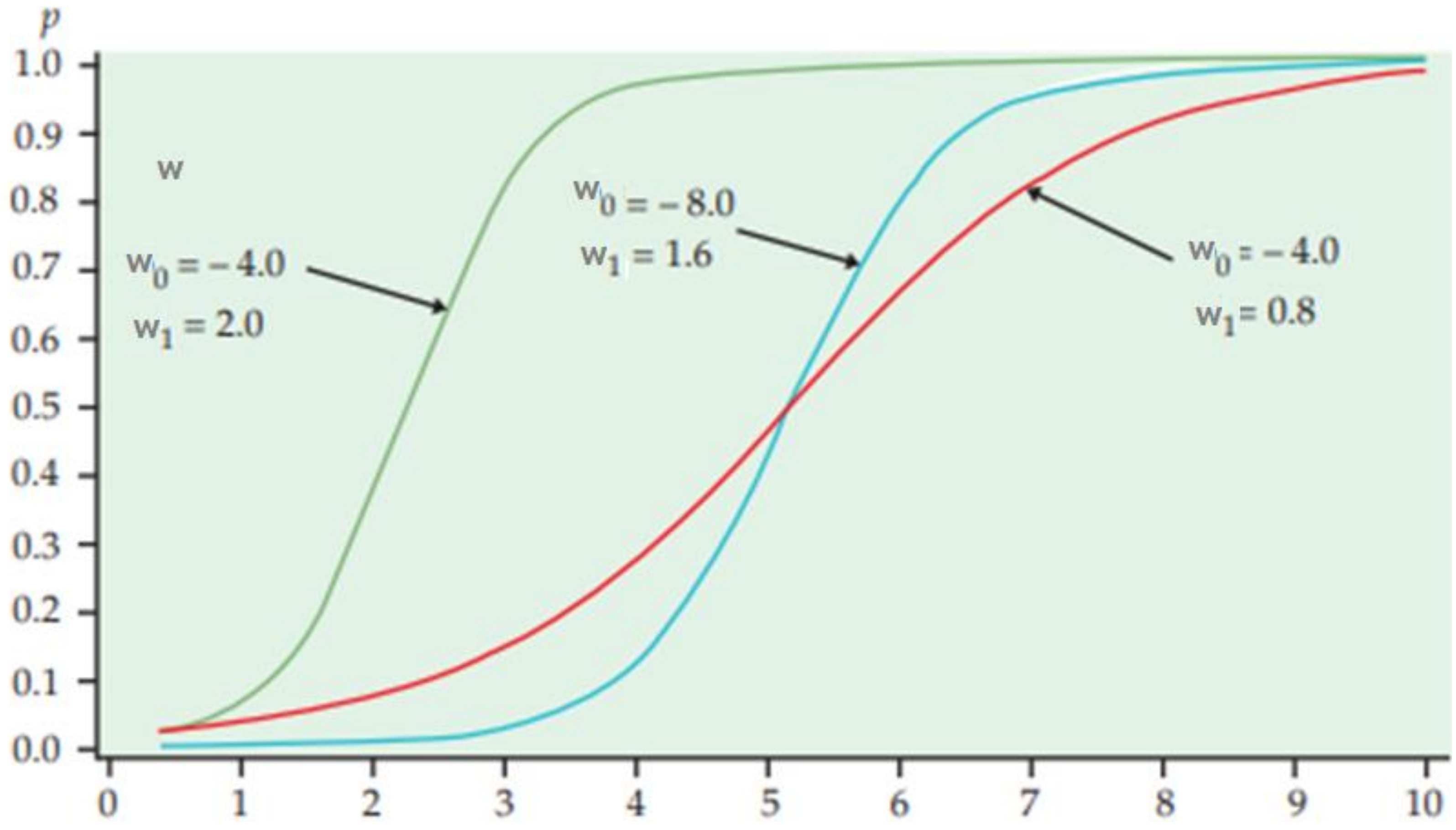
$$y = w_0 + w_1 x = W^T x$$

- Logistic Regression uses Sigmoid Function to model the relationship between input variable and output response.

$$p(x) = W^T x$$

$$\log\left(\frac{p(x)}{1-p(x)}\right) = W^T x = w_0 + w_1 x$$

LOGISTIC REGRESSION



LOGISTIC REGRESSION

$$\mathrm{p}(x) = h_{\boldsymbol{w}}(\mathbf{x}) = \sigma(\mathbf{x}^{\top}\boldsymbol{W})$$

- σ(·) is a sigmoid function
- Outputs a number between 0 and 1
- Logistic Regression model Prediction

$$\hat{y} = \begin{cases} 0 & \text{if } \mathrm{p}(\mathrm{x}) < 0.5 \\ 1 & \text{if } \mathrm{p}(\mathrm{x}) \geq 0.5 \end{cases}$$

LOGISTIC REGRESSION

- Goal: w_0 and w_1 must be estimated.
- Linear Regression uses Least Squared method
- Logistic Regression uses Maximum Likelihood estimation (MLE)
  - For a Binary classification:
  - M labeled samples with labels (0 or 1)
  - For class-1: Find values of W such that p(x) is close to 1
  - For class-0: Find values of W such that p(x) is close to 0 or 1-p(x) is close to 1

LOGISTIC REGRESSION

$$\mathrm{p}(x) = h_{\boldsymbol{w}}(\mathbf{x}) = \sigma(\mathbf{x}^\top \boldsymbol{W})$$

$$\mathrm{p}(x) = \sigma(w_0 + w_1 x_1 + \cdots + w_n x_n)$$

- *Cost function:*

$$Loss(\boldsymbol{W}, \boldsymbol{x}) = \begin{cases} -\log(\mathrm{p}(\mathrm{x})) & \text{if } y = 1 \\ -\log(1 - \mathrm{p}(\mathrm{x})) & \text{if } y = 0 \end{cases}$$

- The *log loss* can be written as:

$$J(W) = -\frac{1}{m} \sum_{i=1}^{m} [y_i \log(p(x_i)) + (1 - y_i) \log(1 - p(x_i))]$$

LOGISTIC REGRESSION

- This cost function is convex
  - Gradient Descent is guaranteed to find the global minimum
- The weights can be updated using the partial derivative of the cost function according to W_i

$$w_i = w_i - \lambda \frac{\partial J}{\partial w_i}$$

$$\frac{\partial}{\partial w_j} \mathrm{J}(\boldsymbol{w}) = \frac{1}{m} \sum_{i=1}^{m} \left(\sigma(\boldsymbol{W}^{\top} x_i) - y_i\right) x_j$$

LOGISTIC REGRESSION

- Logistic regression will find W such that it minimizes J(W)
- Make prediction using

$$p(x) = \frac{1}{1+e^{-W^T X}}$$

$$\hat{y} = \begin{cases} 0 & \text{if } p(x) < 0.5 \\ 1 & \text{if } p(x) \geq 0.5 \end{cases}$$

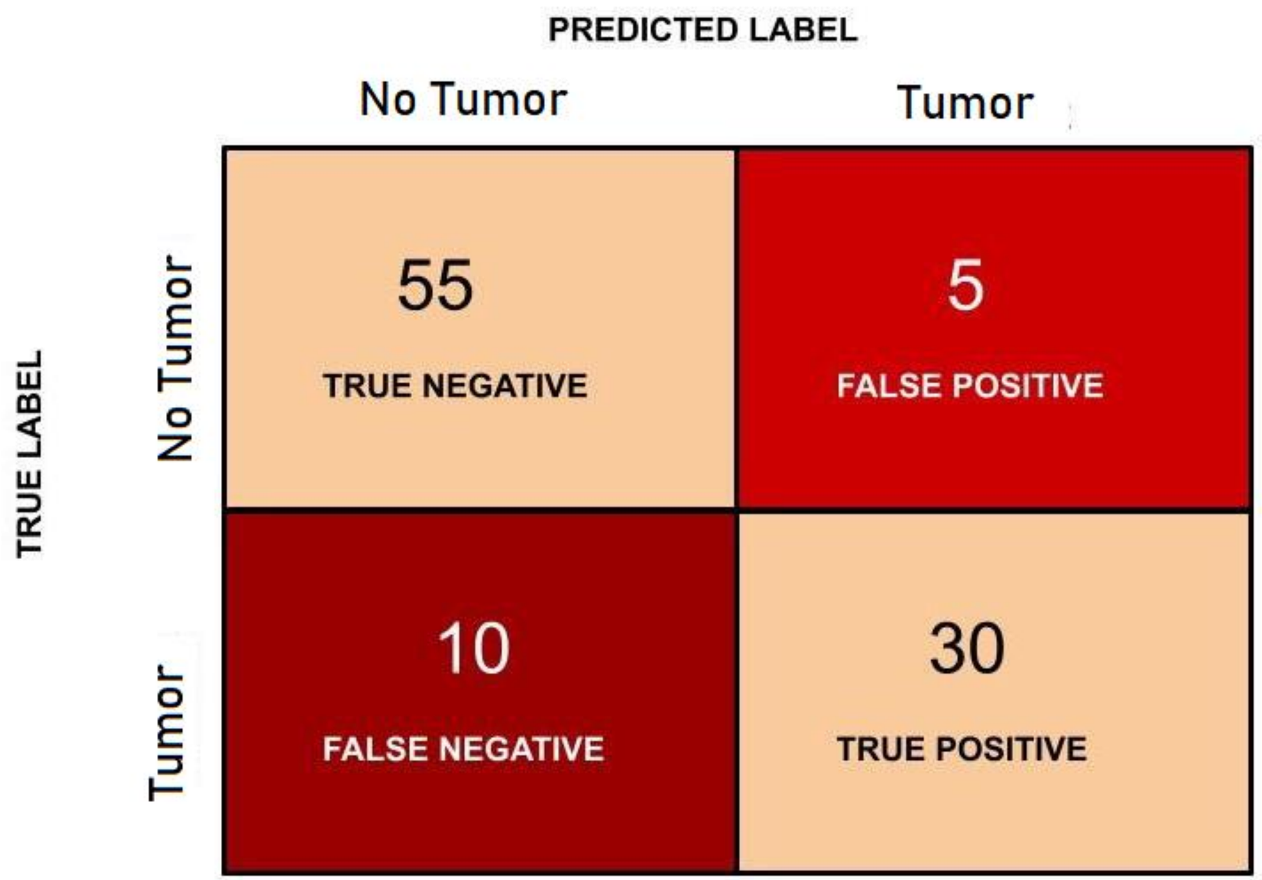
MULTINOMIAL REGRESSION

- Can we extend the Logistic Regression model for multiclass classification(more than two) problems?
  - Yes ☺
- Multinomial model regression
  - We can use One-vs-Rest (One-vs-all)
  - Divide the problem into many sub problems (binary)
  - Train separate model for one class vs rest classes
  - Repeat for all possible combinations for every class

QUANTITATIVE EVALUATION OF ML MODELS.

- Different metrics can be used to produce the quantitative results for a model.
- Classification:
  - We commonly use Precision, Recall, f-Measure, Accuracy.
- Regression
  - R Square, Mean Square Error(MSE), Root Mean Square Error(RMSE), Mean Absolute Error(MAE)

QUANTITATIVE EVALUATION OF ML MODELS.



QUANTITATIVE EVALUATION OF ML MODELS.

A **true positive (TP)** is an outcome where the model *correctly* predicts the *positive* class.

True negative (TN) is an outcome where the model *correctly* predicts the *negative* class.

A **false positive (FP)** is an outcome where the model *incorrectly* predicts the *positive* class.

False negative (FN) is an outcome where the model *incorrectly* predicts the *negative* class.

QUANTITATIVE EVALUATION OF ML MODELS.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$\text{F1} = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

REFERENCE

- Chapter 5, Deep Learning MIT Press 2016, Ian Goodfellow
- Chapter 3 Pattern Recognition and Machine Learning, Christopher M. Bishop
- Some graphics from the internet:
  - https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35

Thank You ☺