# Logistic Regression

Dr. Kalidas Y., IIT Tirupati

In this class, you will learn about logistic regression

# 68) key phrase… “Logistic Regression”

KALIDAS PERSONAL COPY - NOT FOR DISTRIBUTION - 978-0-444-64211-0.pdf - Adobe Reader

via a weight vector. As against a linear regression where $w \cdot x$ is directly used to predict $y$ coordinate, in the logistic regression formulation $w \cdot x$ is defined as log odds in favor of predicted class being 1. But for the interpretation of the $w \cdot x$ values, the rest of the formulation depends on regression of $w$ vector to minimize a *logit* loss function, a name of logistic regression is given for this method. It is essentially a classification algorithm though the word “regression” is present.

Let $y' \in 0, 1$ be the actual label of a data point in a two class problem. Let $D = \{(x', y')\}$ be the data set of points. Let $(\forall (x', y') \in D) : D_1 = \{(x', 1) | y' = 1\}$ and $D_0 = \{(x', 0) | y' = 0\}$ Let $y$ denote the predicted class. Now, let us define $w \cdot x = log(\frac{P(y=1|w,x)}{P(y=0|w,x)})$ where $P(y = 1|w, x)$ denote the probability of predicted class 1 given $w$ and $x$ vectors. With this setting the logistic regression weight update equations are derived as below.

$$P(y=0|w,x) = 1 - P(y=1|w,x)$$

$$w \cdot x = log\left(\frac{P(y=1|w,x)}{P(y=0|w,x)}\right)$$

$$\Rightarrow P(y=1|w,x) = \frac{1}{1+e^{-w \cdot x}}$$

$$P(D|w) = P(D_1|w) \times P(D_0|w)(i.i.d)$$

$$= \prod P(y=1|w,x_1) \times \prod P(y=0|w,x_0)$$

# Logistic Regression Formulation

- $x_i$ is a $k \times 1$ dimensional vector
- $y_i \in \{-1,1\}$
- Model, $y = f(x) = w \cdot x$
  - $w$ is a $k \times 1$ dimensional vector
- $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$
- ??? Loss function $L(w)$

# Logistic Regression Formulation

- $x_i$ is a $k \times 1$ dimensional vector
- $y_i \in \{-1,1\}$
- Model, $y = f(x) = w \cdot x$
  - $w$ is a $k \times 1$ dimensional vector
- $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$
- ??? Loss function $L(w)$
- Odds in favour of +1
  - Probability of prediction of +1 given w and x
  - $P(y = +1 | w, x_i)$

# Logistic Regression Formulation

- $x_i$ is a $k \times 1$ dimensional vector
- $y_i \in \{-1,1\}$
- Model, $y = f(x) = w \cdot x$
  - $w$ is a $k \times 1$ dimensional vector
- $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$
- ??? Loss function $L(w)$

Odds in favour of +1

Probability of prediction of +1 given w and x

$$P(y = +1|w, x_i)$$

Ratio of odds in favour of +1 to -1

$$\frac{P(y = +1|w, x_i)}{P(y = -1|w, x_i)}$$

Define the logarithm of the above ratio as $w \cdot x_i$

$$w \cdot x_i = \log(\frac{P(y = +1|w, x_i)}{P(y = -1|w, x_i)})$$

# Deriving.. $P(y = y_i|w, x_i) = \frac{1}{1+e^{-y_i \times (w \cdot x_i)}}$

- $\log\left(\frac{P(y=+1)}{P(y=-1)}\right) = w \cdot x$
- $\frac{P(y=+1)}{P(y=-1)} = e^{w \cdot x}$
- $P(y = +1) = P(y = -1)e^{w \cdot x}$
- $P(y = +1) = \left(1 - P(y = +1)\right)e^{w \cdot x}$
- $P(y=+1)+P(y=+1)e^{w.x} = e^{w.x}$
- $P(y = +1) = \frac{e^{w \cdot x}}{1+e^{w \cdot x}} = \frac{1}{1+e^{-w \cdot x}}$
- *Similarly we can observe,* $P(y = -1) = 1 - P(y = +1) = \frac{1}{1+e^{w \cdot x}}$
- We can simplify and write, $P(y = y_i|w, x_i) = \frac{1}{1+e^{-y_i \times (w \cdot x_i)}}$

# Probability of Data..
# Minimize $\sum_{i=1}^{i=N} \log(1 + e^{-y_i \times w \cdot x_i})$

- Data set $D = \{(x_1, y_1), \dots, (x_N, y_N)\}$
- Probability of correct predictions
  - on entire data, $P(y = y_1|x_1) \times \cdots P(y = y_N|x_N)$
  - $= \prod_{i=1}^{i=N} P(y = y_i|x_i)$
- Entire probability of data set, $P(D|w) = \prod_{i=1}^{i=N} P(y = y_i|x_i)$
- Logarithm of the probability of data set, $\log\big(P(D|w)\big)$
- $\log\big(P(D|w)\big) = \sum_{i=1}^{i=N} \log(P(y = y_i|x_i))$
- $\sum_{i=1}^{i=N} \log(P(y = y_i|x_i)) = \sum_{i=1}^{i=N} \log(\frac{1}{1+e^{-y_i \times w \cdot x_i}})$
- $= -\sum_{i=1}^{i=N} \log(1 + e^{-y_i \times w \cdot x_i})$
- Maximizing $-\sum_{i=1}^{i=N} \log(1 + e^{-y_i \times w \cdot x_i})$ ***is equivalent to*** Minimizing $\sum_{i=1}^{i=N} \log(1 + e^{-y_i \times w \cdot x_i})$

Maximizing ***–f(x)*** becomes Minimization ***+f(x)***

Look at this minus symbol?

(C) Dr. Kalidas Y., IIT Tirupati