CPE/EE 695: Applied Machine Learning

Lecture 3 - 3: Logistic Regression

Dr. Shucheng Yu, Associate Professor
Department of Electrical and Computer Engineering
Stevens Institute of Technology

Logistic Regression

Usually be used for binary **classification:**

Pr(Y|X), where Y is a binary variable. (why probability?)

Logistic Regression

Usually be used for binary **classification:**

Pr(Y|X), where Y is a binary variable. (why probability?)

E.g., Pr(Tomorrow snow | today windy)

Logistic Regression

Usually be used for binary **classification:**

Pr(Y|X), where Y is a binary variable. (why probability?)

E.g., Pr(Tomorrow snow | today windy)

Let Pr(Y=1 | X=x) = $p(x;\theta)$

Logistic Regression

Usually be used for binary **classification:**

Pr(Y|X), where Y is a binary variable. (why probability?)

E.g., Pr(Tomorrow snow | today windy)

Let Pr(Y=1 | X=x) = $p(x;\theta)$

Maximize likelihood:

$$\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

Logistic Regression

Usually be used for binary **classification:**

Pr(Y|X), where Y is a binary variable. (why probability?)

E.g., Pr(Tomorrow snow | today windy)

Let Pr(Y=1 | X=x) = $p(x;\theta)$

Assumption: p is modeled with parameter θ; otherwise, optimization problem doesn't work

Maximize likelihood:

$$\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

Logistic Regression

Let Pr(Y=1 | X=x) = $p(x;\theta)$

Maximize likelihood:

$$\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

Logistic Regression

Let Pr(Y=1 | X=x) = $p(x;\theta)$

Maximize likelihood:

$$\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

Task: to estimate θ by maximizing the likelihood.

Logistic Regression

Let Pr(Y=1 | X=x) = $p(x;\theta)$

Maximize likelihood:

$$\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

Task: to estimate θ by maximizing the likelihood.

How can we use **linear regression** to solve this?

Logistic Regression

Let Pr(Y=1 | X=x) = $p(x;\theta)$

Task: to estimate θ by maximizing the likelihood.

How can we use **linear regression** to solve this?

Attempt 1: assume $p(x;\theta)$ be a linear function of x

Attempt 2: assume $\log p(x;\theta)$ be a linear function of x

Attempt 2: assume $\log \frac{p}{1-p}$ be a linear function of x

Remember: 0<= p <= 1

Logistic Regression

Logistic regression model

$$\log \frac{p}{1-p} = \beta_0 + x \cdot \beta$$

Which gives $p(x) = \frac{1}{1+ e^{-\beta_0 + x \cdot \beta}}$

Logistic Regression

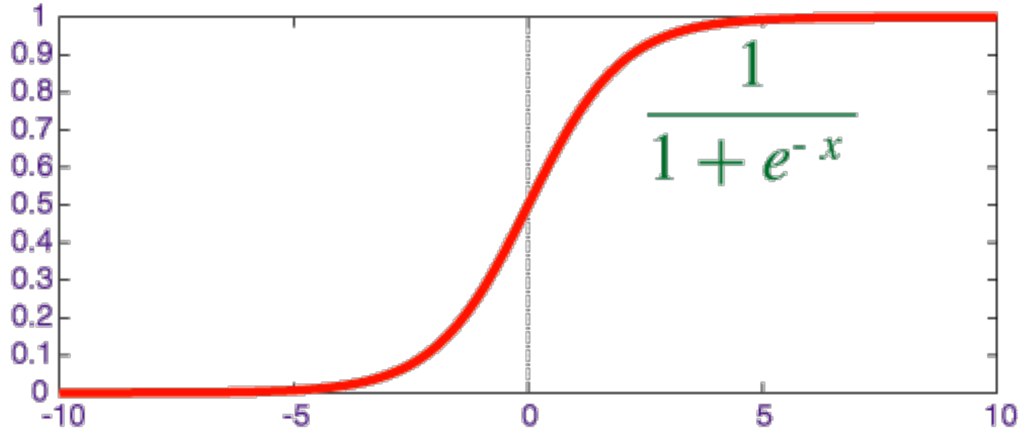
Logistic Regression model estimated probability:

$$\hat{p} = h_w(x) = \sigma(w^T \cdot x)$$

where $\sigma(\cdot)$ is a logistic function (or sigmoid function).

Prediction:

$$\hat{y} = \begin{cases} 0, & if\ \hat{p} < 0.5 \\ 1, & if\ \hat{p} \geq 0.5 \end{cases}$$



Logistic Regression

Logistic regression cost function (log loss):

$$\mathrm{J}(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(\hat{p}^{(i)}) + \left(1 - y^{(i)}\right)\log(1 - \hat{p}^{(i)}))]$$

Cost for a single training instance (encourage positive instances but discourage negative instances):

$$\mathrm{c}(\theta) = \begin{cases} -\log(\hat{p})\,, if\ y = 1 \\ -\log(1 - \hat{p}), if\ y = 0 \end{cases}$$

Logistic Regression

Training in logistic regression

No Normal Equation (i.e., closed form solution) for w.

But the cost function is convex and derivable. **Gradient Descent** is guaranteed to find global maximum.

Logistic cost function partial derivatives:

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (\delta\left(\theta^T \cdot x^{(i)}\right) - y^{(i)}) x_j{}^{(i)}$$

Logistic Regression

Muli-class classification:

1) Multiple binary classifier (one vs. the rest)

Logistic Regression

Muli-class classification:

2) Softmax Regression (or Multinomial Logistics Regression)

i) first compute a score $s_k(x) = \ {\theta_k}^T \cdot x$

ii) then compute softmax function:

$$\hat{p}_k = \delta(s_k(x))_k = \frac{\exp(s_k(x))}{\sum_{j=1}^{K} \exp(s_j(x))}$$

Training: to minimize the **cross-entropy** cost function:

$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} {y_k}^{(i)} \log({\hat{p}_k}^{(i)})$$

${y_k}^{(i)}$ = 1 if the ith instance belongs to class k; 0 otherwise.

Logistic Regression (cont')

Muli-class classification:

2) Softmax Regression (or Multinomial Logistics Regression)

The gradient for **cross-entropy** cost function:

$$\nabla_{\theta_k} J(\Theta) = \frac{1}{m}\sum_{i=1}^{m}\left({\hat{p}_k}^{(i)} - {y_k}^{(i)}\right)x^{(i)}$$

${y_k}^{(i)}$ = 1 if the ith instance belongs to class k; 0 otherwise.

STEVENS

INSTITUTE of TECHNOLOGY

THE INNOVATION UNIVERSITY®

stevens.edu