# Linear Regression and Logistic Regression

Khanh Nguyen

June 02, 2015

We consider data with response $y$ and input $x = \{x_1, x_2, ..., x_D\}$. We model the relationship between $x$ and $y$ by a probabilistic model of $p(y|x)$.

Linear regression and logistic regression are examples of parametric models, i.e. models that have fixed amounts of parameters. We denote by $w$ the model parameter.

## 1 Linear Regression

In linear regression, $y$ is a linear function of $x$:

$$y(x) = w^T x + \epsilon = \sum_{j=1}^{D} w_j x_j + \epsilon \tag{1}$$

We assume that $\epsilon$ has a Gaussian distribution, i.e. $\epsilon \sim \mathcal{N}(w_0, \sigma^2)$. We can rewrite the linear regression model as a conditional probability density:

$$p(y|x, \theta) = \mathcal{N}(y|w^T x, \sigma^2) \tag{2}$$

where $\theta = (w, \sigma^2)$ [1]. $\mathcal{N}(y|w^T x, \sigma^2)$ is a normal distribution with mean depending on $x$. To represent non-linear relationship between $x$ and $y$, we can replace the mean and variance of the distribution by any non-linear function of $x$.

## 2 Logistic Regression

Despite of the name, logistic regression is used for classification. $y$ is now a binary variable, taking value of either 0 or 1. Hence, we need a different distribution than Gaussian to model this fact. Bernoulli distribution is a natural choice:

$$p(y|x, w) = Ber(y|\mu(x)) \tag{3}$$

where $\mu(x)$ is the parameter for the Bernoulli distribution and is a function of $x$.

Since $\mu(x)$ must be between 0 and 1, we cannot use $w^T x$ but have to transform it somehow to fit into the [0,1] interval. We use the sigmoid function, which is defined as:

[1] Note that here $x$ and $w$ are slightly different from Eqn. 1: $x = \{1, x1, ..., x_D\}$ and $w = \{w_0, w_1, .., w_D\}$

$$sigm(t) = \frac{1}{1 + e^{-t}} = \frac{e^t}{1 + e^t} \tag{4}$$

Setting $\mu(x) = sigm(x)$, we obtain the logistic regression model:

$$p(y|x, w) = Ber(y|sigm(x)) \tag{5}$$