CS4801 : Logistic Regression

Sahely Bhadra

23/8/2017

1. Logistic Regression : probabilistic inter pretation
2. Loss function for LR
3. LR is a linear classifier
4. Gradient descent algorithm
5. LR for multi class classification

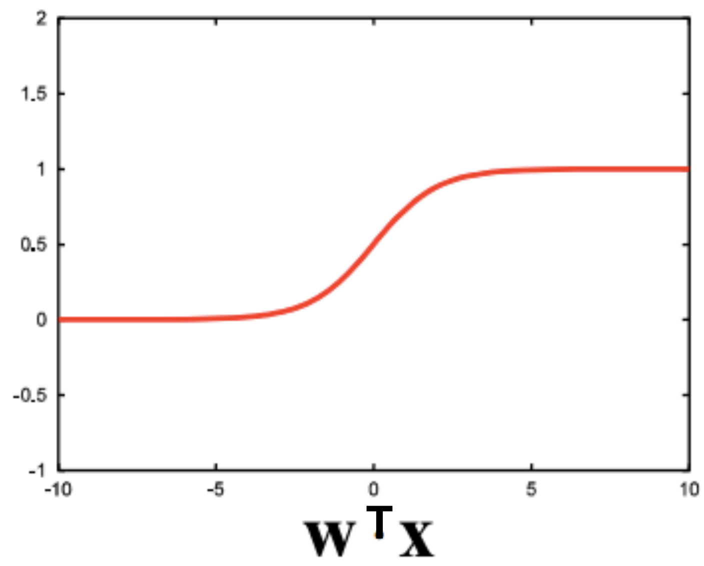
Logistic Regression

- Posterior is defined as

$$p(y=1\,|\,\mathbf{x};\mathbf{w}) = g(\mathbf{x},\mathbf{w}) = \frac{1}{1+e^{-\mathbf{w}^\mathsf{T}\mathbf{x}}}$$

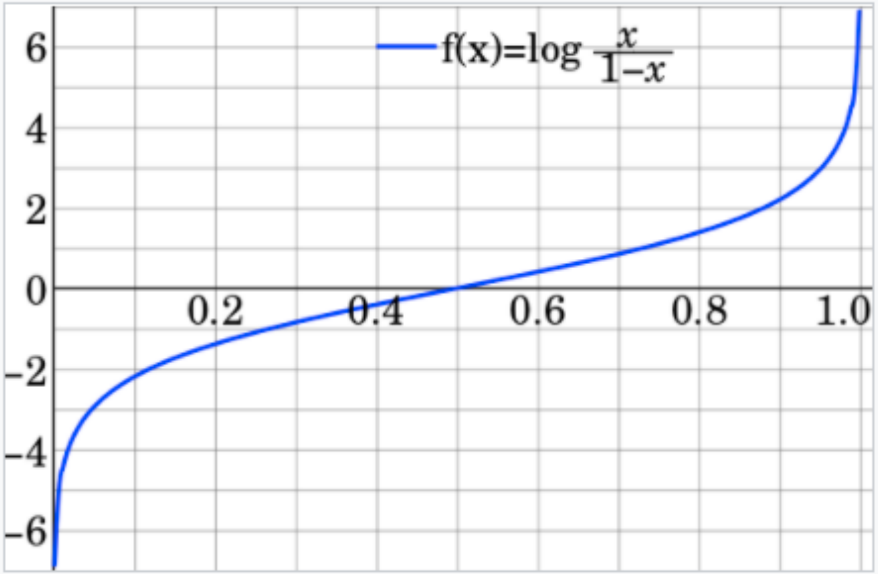
$$p(y=0\,|\,\mathbf{x};\mathbf{w}) = 1 - g(\mathbf{x},\mathbf{w})$$

$$g(\mathbf{x},\mathbf{w}) = \frac{1}{1+\exp(-\mathbf{w}^\mathsf{T}\mathbf{x})}$$



Logistic Regression

$$\operatorname{logit}(p) = \log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p) = -\log\left(\frac{1}{p} - 1\right).$$



Plot of logit(p) in the domain of 0 to 1, where the base of logarithm is e

$$\operatorname{logit}^{-1}(\alpha) = \operatorname{logistic}(\alpha) = \frac{1}{1+\exp(-\alpha)} = \frac{\exp(\alpha)}{\exp(\alpha)+1}$$

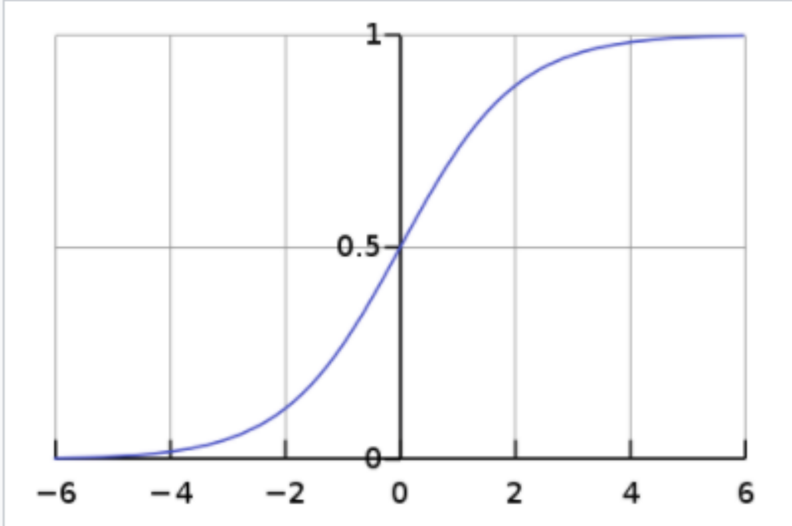


Figure 1. The standard logistic function $\sigma(t)$; note that $\sigma(t) \in (0,1)$ for all t.

Logistic Regression

Logistic Regression : parametric assumption for posterior distribution

$$P(y=C_1/x) = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}$$

$$P(y=C_2/x) = 1 - P(y=C_1/x)$$

Hence assuming $y=C_1 \Rightarrow y=+1$ and $y=C_2 \Rightarrow y=-1$

$$P(y/x) = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}.y}}$$

Logistic Regression : Loss function

Logistic Regression : parametric assumption for posterior distribution

Hence assuming y=C_1=>y=+1 and y=C_2=>y=-1

$$P(y/x) = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}.y}}$$

By maximising log P(y/x):

$$\min_{\mathbf{w}\in\mathbb{R}^d} \sum_{i}^{N} \underbrace{\log\left(1+e^{-y_i f(\mathbf{x}_i)}\right)}_{\text{loss function}} + \lambda \underbrace{||\mathbf{w}||}_{\text{regularization}}{}^2$$

Can be solved using Gradient Descent

Logistic Regression : Linear Classifier

P(y=1/x) = $\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}$

Hence we predict y=1 if $\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}$ >= 0.5

or 1 >= $e^{-\mathbf{w}^T\mathbf{x}}$

or $w^Tx >= 0$

Hence Logistic regression is a Linear Function

Multi classes case

Choose class K to be the “reference class” and represent each of the other classes as a logistic function of the odds of class *k* versus class K:

$$\log \frac{P(y=1 \mid \mathbf{x})}{P(y=K \mid \mathbf{x})} = \mathbf{w}_1{}^{\mathsf{T}}\mathbf{x}$$

$$\log \frac{P(y=2 \mid \mathbf{x})}{P(y=K \mid \mathbf{x})} = \mathbf{w}_2{}^{\mathsf{T}}\mathbf{x}$$

$$\vdots$$

$$\log \frac{P(y=K-1 \mid \mathbf{x})}{P(y=K \mid \mathbf{x})} = \mathbf{w}_{K-1}{}^{\mathsf{T}}\mathbf{x}$$

$$P(y=k \mid \mathbf{x}) = \frac{\exp(\mathbf{w}_k{}^{\mathsf{T}}\mathbf{x})}{1+\sum_{l=1}^{K-1}\exp(\mathbf{w}_l{}^{\mathsf{T}}\mathbf{x})} \qquad P(y=K \mid \mathbf{x}) = \frac{1}{1+\sum_{l=1}^{K-1}\exp(\mathbf{w}_l{}^{\mathsf{T}}\mathbf{x})}$$