Machine Learning, Tutorial 3
University of Bern

Mehdi Noroozi (noroozi@inf.unibe.ch)

10/10/2018

Logistic Regression

1. Compute the Hessian of the logistic regression training optimisation problem.

   Solution

   - The function we want to maximise is :
   $$\log L(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] \tag{1}$$
   where $h_\theta(x^{(i)}) = g(\theta^T x^{(i)})$ and $g(x) = \frac{1}{1+e^{-x}}$. Remember that $x^{(i)} \in \mathbf{R}^{n+1}$ and $\theta \in \mathbf{R}^{n+1}$, while $h_\theta(x^{(i)})$ and $y^{(i)}$ are scalar. m is the number of training samples and $x^{(i)}$ and $y^{(i)}$ are the samples i in the training set.
   - the gradient of $\log L(\theta)$ is:
   $$\nabla \log L(\theta) = \frac{1}{m}\sum_{i=1}^{m}(y - h_\theta(x^{(i)}))x^{(i)} \tag{2}$$
   - the Hessian of $\log L(\theta)$ is:
   $$\nabla^2 \log L(\theta) = -\frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))x^{(i)}(x^{(i)})^T \tag{3}$$

Generalized Linear Models

1. Consider the Laplace distribution
   $$p(y;\lambda) = \frac{1}{\lambda}\exp\left(-\frac{|y-\mu|}{\lambda}\right) \tag{4}$$
   where μ is a constant. Is it part of the exponential family of distributions?
2. If we parametrize the Laplace distribution on μ, is it still part of the exponential distributions?

   Solution The class of exponential distributions is defined as all the distributions that can be written in the form
   $$p(y;\eta) = b(y)\exp(\eta^T T(y) - a(\eta)) \tag{5}$$
   where η is called natural parameter, $T(y)$ is called sufficient statistic and $a(\eta)$ is called log partition function.

   The Laplace distribution, if parametrized only on λ, with μ fixed, it is indeed possible to write in the above form:
   $$p(y;\lambda) = \frac{1}{\lambda}\exp\left(-\frac{|y-\mu|}{\lambda}\right) \tag{6}$$
   that is, $\eta = -\frac{1}{\lambda}$, $b(y) = 1$, $T(y) = |y-\mu|$ and $a(\eta) = \log\left(-\frac{1}{\eta}\right) = \log(\lambda)$.

   However, if we parametrize the Laplace distribution on μ, keeping λ fixed, it is not possible to write the distribution in the above form, and it is indeed not part of the exponential family.