Machine Learning
exercise 3

Marc Toussaint

Machine Learning & Robotics lab, FU Berlin

Arnimallee 7, 14195 Berlin, Germany

April 27, 2011

1   Log-likelihood gradient and Hessian

Consider a binary classification problem with data $D = \{(x_i, y_i)\}_{i=1}^n$, $x_i \in \mathbb{R}^d$ and $y_i \in \{0, 1\}$. We define

$$f(x) = x^\top \beta \tag{1}$$

$$p(x) = \sigma(f(x)), \quad \sigma(z) = 1/(1 + e^{-z}) \tag{2}$$

$$L(\beta) = \sum_{i=1}^n \left[y_i \log p(x_i) + (1 - y_i) \log[1 - p(x_i)] \right] \tag{3}$$

where $\beta \in \mathbb{R}^d$ is a vector. (NOTE: the $p(x)$ we defined here is a short-hand for $p(y = 1|x)$ on slide 03:10.)

a) Compute the derivative $\frac{\partial}{\partial \beta} L(\beta)$. Tip: use the fact $\frac{\partial}{\partial z}\sigma(z) = \sigma(z)(1 - \sigma(z))$.

b) Compute the 2nd derivative $\frac{\partial^2}{\partial \beta^2} L(\beta)$.

2   Logistic Regression

On the course webpage there is a data set data2Class.txt for a binary classification problem. Each line contains a data entry (x, y) with $x \in \mathbb{R}^2$ and $y \in \{0, 1\}$.

a) Compute the optimal parameters β and the log-likelihood ($\frac{1}{n} L(\beta)$) of logistic regression using linear features. Plot the probability $p(y = 1 | x)$ over a 2D grid of test points.

Useful gnuplot commands:

```
splot [-2:3][-2:3][-3:3.5] 'model' matrix us ($1/20-2):($2/20-2):3 with lines notitle
plot [-2:3][-2:3] 'data2Class.txt' us 1:2:3 with points pt 2 lc variable title 'train'
```

b) Compute and plot the same for quadratic features.