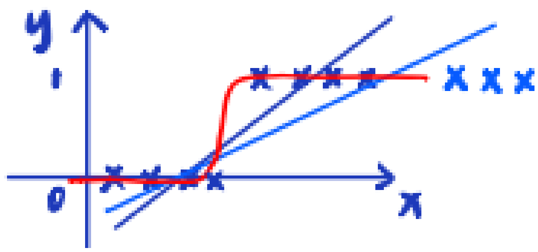
HW7-Notes

Yiping Lu and Jody Shu

March 2020

1 Softmax Regression Algorithm

Softmax is an extension of Logistic Regression (LR).
LR setup : $y \in \{0, 1\} \leftarrow discrete$
want: $h_\theta \in [0, 1]$



Linear reg: [-∞, ∞]
Choose: $h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$,
where $g(z) = \frac{1}{1+e^{-z}}$
(note: $z \to -\infty$ $g(z), \to 0$,
$z \to \infty$ $g(z), \to 1$)
g is called sigmoid/Logistic function
Here: g:$(-\infty, \infty) \to (0, 1)$
p(y=1| $x; \theta) = h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$
p(y=0| $y; \theta) = 1 - h_\theta(x) \geq 0$ $(0 \leq h_\theta(x) \leq 1)$
Next, we are going to write the prob. fcn into one equation.

$p(y \mid x;\theta) = h_\theta(x)^y(1 - h_\theta(x))^{1-y}$
$y = 1 \Rightarrow p(y = 1 \mid x;\theta) = h_\theta(x)^1(1 - h_\theta(x))^0$
$y = 0 \Rightarrow p(y = 0 \mid x;\theta) = 1 - h_\theta(x)$
$L(\theta) = p(y = 0 \mid x;\theta) = \prod_i p(y^i \mid x^i;\theta) = \prod_i h_\theta(x^i)^{y^i}(1 - h_\theta(x^i))^{1-y^i})$
It is much easier to maximize the log likelihood.
$l(\theta) = logL(\theta) = \sum_i [y^i log h_\theta(x^i) + (1 - y^i) log(1 - h_\theta(x^i))]$
How to maximize it? Use (stochastic) gradient descent.
$\theta := \theta \bigoplus \alpha \nabla_\theta l(\theta)$
$\bigoplus$: along with the gradient direction.
$\frac{\partial}{\partial \theta_j} l(\theta) = \sum_{i=1}^n (y^i - h_\theta(x^i))x_j^i$