Logistic Regression

Logistic Regression is a simple statistical model for binary classification (0,1). Given some attributes we want to classify whether a sample should be classified as 0 or 1. Unlike, linear regression, the output is discrete.

Model

[OUTPUT] $\rightarrow$ {0,1}
$\rightarrow z = X\vec{w} + b$
$\Theta \rightarrow (\vec{w}, b)$
[HYPOTHESIS] $h(\Theta, X) = \sigma(z)$

Cost function

$$J(\Theta) = -y\log(h(\Theta, x)) - (1-y)\log(1 - h(\Theta, x))$$

We want the loss function to be convex in $\vec{w}, b$

We know that $log(1 - \frac{1}{1+e^{-z}})$ and $log(\frac{1}{1+e^{-z}})$ are both concave in z
$\therefore$ the loss function is convex in Θ

It is important for loss function to be convex so that we get a global optima and not just local optima or saddle points.

Gradient Descent

Note that we often write $\frac{\partial x}{\partial y}$ as just ∂x

$$\vec{w} := \vec{w} - \alpha\partial J$$

$$b := b - \alpha\partial J$$

We can calculate the gradients as

$$\frac{\partial J}{\partial w} = \frac{\partial J}{\partial z}\frac{\partial z}{\partial w}$$

$$\frac{\partial J}{\partial w} = X^T\left(\frac{1}{1+e^{-z}} - y\right)\frac{1}{n}$$

$$\frac{\partial J}{\partial b} = \left(\frac{1}{1+e^{-z}} - y\right)$$

Questions

1. Can you use it for more than just 2 classes?
2. Should you use Stochastic Gradient descent? if Yes, when?
3. What is ordinal logistic regression?
4. How is ordinal logistic regression different from logistic regression with just more than 2 categories?
5. Explain the need of regularization.