10. Gradient Descent for logistic regression.

I. Training logistic regression.

Find $\vec{w}, b$.

Given new $\vec{x}$, output $f_{\vec{w},b}(\vec{x}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$

which is $P(y=1 \mid \vec{x}; \vec{w}, b)$

II. Gradient Descent. for logistic regression.

Now, our cost looks like this.

$$J(\vec{w},b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) + (1-y^{(i)})\log\left(1-f_{\vec{w},b}(\vec{x}^{(i)})\right)\right]$$

repeat $\{$ $w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{w},b)$

$b = b - \alpha \frac{\partial}{\partial b} J(\vec{w},b)$ $\}$ Now, $j = 1, \ldots, n$.

simultaneous updates.

Now, $\frac{\partial}{\partial w_j} J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right) \cdot x_j^{(i)}$

j^{th} feature.

$\frac{\partial}{\partial b} J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)$

repeat $\{$ $w_j = w_j - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right) \cdot x_j^{(i)}\right]$

$b = b - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)\right]$

$\}$ simultaneous updates.

looks like linear regression, but definition of f is changed.

Linear regression $f_{\vec{w},b}(\vec{x}) = \vec{w}\cdot\vec{x} + b$

Logistic regression $f_{\vec{w},b}(\vec{x}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$

III. Same Concepts.

a. Monitor gradient descent (learning curve)
b. Vectorized implementation.
c. feature scaling.