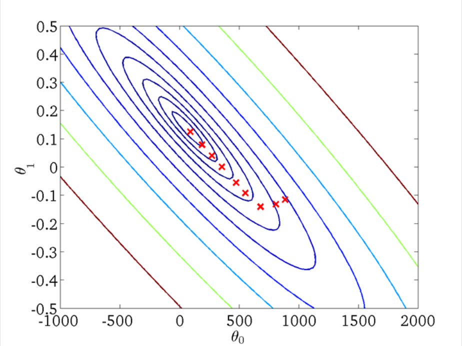
Batch vs Stohastic descent

Repeat until convergence {

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)} \text{ (for each j) .}$$

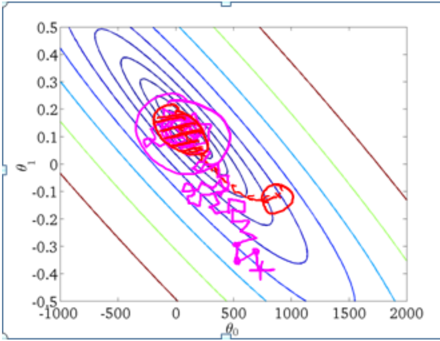
}



Repeat until convergence {
for i = 1 to m

$$\theta_j = \theta_j - \alpha(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)} \text{ (for each j) .}$$

}



Regularization

March 12, 2013

Bayesian statistics and regularization

- Recently we viewed θ as an unknown parameter and estimate it using maximum likelihood

$$\theta_{ML} = \text{argmax}_{\theta} \prod_{i=1}^{n} p(y^{(i)}|x^{(i)};\theta)$$

- Lets think of θ as being a random variable distributed by some prior distribution $p(\theta)$
- Given a training set $S = \{(x^{(i)}, y^{(i)})\}_{i=1}^{m}$ lets compute posterior

$$p(\theta|S) = \frac{P(S|\theta)P(\theta)}{p(S)} = \frac{(\prod_{i=1}^{m} p(y^{(i)}|x^{(i)},\theta))p(\theta)}{\int_{\theta}(\prod_{i=1}^{m} p(y^{(i)}|x^{(i)},\theta))p(\theta)d\theta}$$

- In general it is very hard to estimate $p(\theta|S)$ over θ
- In practice

$$\theta_{MAP} = \text{argmax}_{\theta} \prod_{i=1}^{m} p(y^{(x_i)}|x^{(y_i)},\theta)p(\theta)$$

- Common choice $\theta\ \mathcal{N}(0, 1/\lambda I)$, the norm of θ usually less then that selected by ML

Regularized Cost Functions

- Least-squares cost function

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{m\lambda}{2}\sum_{j=1}^{n}\theta_j^2$$

- Logistic regression cost function

$$J(\theta) = \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1 - y^{(i)}) \log(1 - h(x^{(i)})) + \frac{m\lambda}{2}\sum_{j=1}^{n}\theta_j^2$$

Generalizaed Linear Models

March 12, 2013

Motivation

So far, we've seen

- Regression

$$p(y|x;\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\theta^T x)^2}{2\sigma^2}\right\} \quad \text{Normal}(\mu, \sigma^2)$$

- Classification

$$p(y|x;\theta) = (h_\theta(x))^y (1-h_\theta(x))^{1-y} \quad \text{Bernoulli}(\phi)$$

- Could we generalize this?

Exponential Family

$$p(y;\eta) = b(y)\exp(\eta^T T(y) - a(\eta))$$

- η - natural parameter (actual parameter of distribution)
- $a(\eta), b(y)$ - specify the specific form of the distribution
- $T(y)$ - might be a vector

Gaussian Distribution

- Let $\sigma = 1$

$$p(y;\mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{1}{2}(y-\mu)^2\right\} =$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} \cdot \exp\left\{\mu y - \frac{1}{2}\mu^2\right\}$$

- $\eta = \mu$
- $T(y) = y$
- $a(\eta) = \mu^2/2 = \eta^2/2$
- $b(y) = (1/\sqrt{2\pi}) \exp\left\{-y^2/2\right\}$

Bernoulli Distribution

$$p(y;\phi) = \phi^y(1-\phi)^{1-y} =$$

$$\exp\{y\log\phi + (1-y)\log(1-\phi)\}$$

$$\exp\left\{\log\left(\frac{\phi}{1-\phi}\right)y + \log(1-\phi)\right\}$$

- $\eta = \log\frac{\phi}{1-\phi}$, $\phi = \frac{1}{1+e^{-\eta}}$
- $T(y) = y$
- $a(\eta) = -\log(1-\phi) = \log(1+e^{\eta})$
- $b(y) = 1$

General Lenear Models

Assume

- $y|x;\theta$ ExpFamily(η)
- Given x, goal is to output $E[T(y)|x]$.
  - Want $h(x) = E[T(y)|x]$
- $\eta = \theta^T x$

Ordinary Least Squares

- Let $y|x$ is Gaussian, then this Exponential distribution with parameter $\eta = \mu$
- So

$$\begin{aligned} h_\theta(x) &= E[y|x;\theta] \\ &= \mu \\ &= \eta \\ &= \theta^T x \end{aligned}$$

Logistic Regression

- Let $y|x$ is Bernoulli, then this Exponential distribution with $\phi = 1/(1 - e^{-\eta})$
- So

$$\begin{aligned} h_\theta(x) &= E[y|x;\theta] \\ &= \phi \\ &= 1/(1 - e^{-\eta}) \\ &= 1/(1 - e^{-\theta^T x}) \end{aligned}$$