Lecture 3 Locally Weighted & Logistic Regression

1. Parametric Learning Algorithms & Non-Parametric Learning Algorithms
   Parametric: Fit fixed set of parameters (θ_i)
   Non-Parametric: Amount of data/parameters keeps growing (linearly) with the size of training set. (Locally Weighted Regression)

2. Locally Weighted Regression
   Fit θ to minimize
   $$\sum_{i=1}^{m} w^{(i)} \left(y^{(i)} - \theta^T x^{(i)}\right)^2$$, where $w^{(i)}$ is a weighted function,
   $$w^{(i)} = \exp\left(-\frac{(x^{(i)} - x)^2}{2\tau^2}\right)$$
   ⇓
   put more weight on the training data whose x is closer to the target x

   ⇒ If $x^{(i)} - x$ is small, $w^{(i)} \approx 1$

   If $x^{(i)} - x$ is large, $w^{(i)} \approx 0$

τ represents the bandwidth

If τ is too big, underfitting; If τ is too small, overfitting.

Appropriate for low-dimensional training data set.

3. Probabilistic Interpretation

Assume $y^{(i)} = \theta^T x^{(i)} + \varepsilon^{(i)}$ → IID (independently & identically distribute)

$\varepsilon^{(i)}$ → error: unmodelled effects, random noise

$$\varepsilon^{(i)} \sim N(0, \sigma^2) \Rightarrow P(\varepsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\varepsilon^{(i)})^2}{2\sigma^2}\right)$$

$$\Rightarrow P(y^{(i)} \mid x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$

I.e. $y^{(i)} \mid x^{(i)}, \theta \sim N(\theta^T x^{(i)}, \sigma^2)$

Likelihood Function: $L(\theta) = L(\theta; X, \vec{y}) = p(\vec{y} \mid X; \theta)$

$$= \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)}; \theta)$$

maximize

$$\Rightarrow \text{Maximum Likelihood} = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$

Log likelihood: $l(\theta) = \log L(\theta)$

<proof> $= \log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$

$= \sum_{i=1}^{m} \left[\log \frac{1}{\sqrt{2\pi}\sigma} + \log \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)\right]$

$= m \log \frac{1}{\sqrt{2\pi}\sigma} + \sum_{i=1}^{m} -\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}$

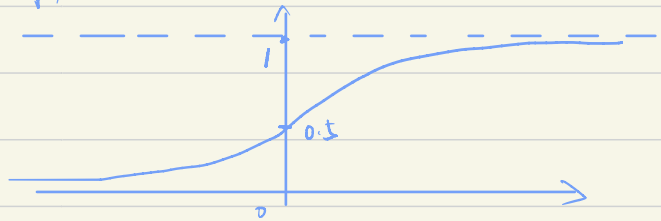
According to maximum likelihood, choose θ to maximize $= \sum_{i=1}^{m} \frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2} = J(\theta)$

least square = choose θ to minimize

4. Logistic Regression (把LR的output变到0,1之间)

logistic function: $g(z) = \frac{1}{1+e^{-z}} \in [0,1]$

$\Rightarrow h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$



$$\begin{cases} P(y=1|x;\theta) = h_\theta(x) \\ P(y=0|x;\theta) = 1 - h_\theta(x) \end{cases} \Rightarrow P(y|x;\theta) = h_\theta(x)^y (1-h_\theta(x))^{1-y}$$

$$\Rightarrow L(\theta) = P(\vec{y}|x;\theta) = \prod_{i=1}^{m} h_\theta(x)^{y^{(i)}} (1-h_\theta(x))^{1-y^{(i)}}$$

$$\Rightarrow \ell(\theta) = \sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log (1-h_\theta(x^{(i)}))$$

Choose θ to maximize $\ell(\theta)$: Batch Gradient Ascend

$$\theta_j := \theta_j + \alpha \frac{\partial}{\partial \theta_j} \ell(\theta) = \theta_j + \alpha \sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)})) x_j^{(i)}$$ equals to gradient descend

5. Newton's Method

Have a f, want to find a θ, s.t. $f(\theta) = 0$

(want to maximize $\ell(\theta) \Rightarrow \ell'(\theta) = 0$)

$\theta^{(1)} := \theta^{(0)} - \boxed{\Delta}$ → 下图

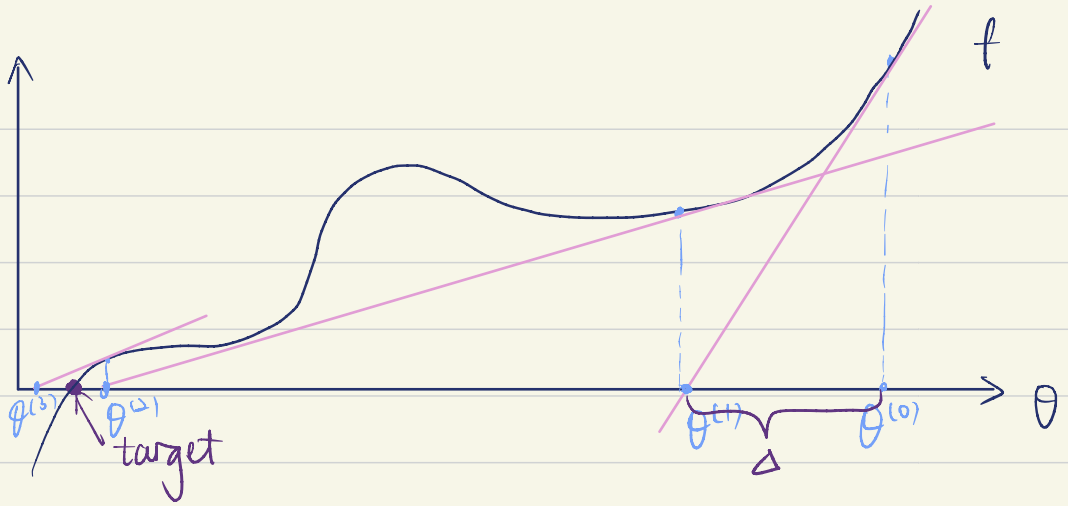
$f'(\theta^{(0)}) = \frac{f(\theta^{(0)})}{\Delta}$

$\Delta = \frac{f(\theta^{(0)})}{f'(\theta^{(0)})}$

$\Rightarrow \theta^{(t+1)} := \theta^{(t)} - \frac{f(\theta^{(t)})}{f'(\theta^{(t)})}$

let $f(\theta) = \ell'(\theta)$

$\Rightarrow \theta^{(t+1)} := \theta^{(t)} - \frac{\ell'(\theta^{(t)})}{\ell''(\theta^{(t)})}$



when θ is a vector, ($\theta \in \mathbb{R}^{n+1}$)

$\theta^{(t+1)} := \theta^{(t)} - H^{-1}\nabla_\theta l$, where H is the Hessian matrix,

$H_{ij} = \frac{\partial^2 l}{\partial\theta_i \cdot \partial\theta_j}$

H^{-1}: $\mathbb{R}^{(n+1)\times(n+1)}$; $\nabla_\theta l$: vector $\mathbb{R}^{n+1}$

Appropriate for the scenario when # parameter is small (50, 100)

↖ more computational