EECS 545

Chien-Wei Lin

1. Probability

(1)

Let event $(H=h) = A$, $(D=d) = B$

Then, by bayes' Rule

$$P(A \mid B) = \frac{P(B|A)}{P(B)} P(A)$$

(i) it depends on the probability of $P(B|A)$ and $P(B)$

if $P(B|A) \geq P(B)$

then $P(H=h \mid D=d) \leq P(H=h)$

otherwise

$P(H=h \mid D=d) > P(H=h)$

(ii)

$$P(H=h) \geq P(D=d \mid H=h) P(H)$$

since $P(B) \leq 1$

(2)

(i) $$E_Y[E_X[X|Y]] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, p_{X|Y=y}(x|y)\, dx\, p(y)\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, p_{Y|X}(y|x)\, dy\, dx = \int_{-\infty}^{\infty} x\, p(x)\, dx = E[X]$$

(ii)

$$Var[X] = E[X^2] - E[X]^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 p(x|y)\, dx\, p(y)\, dy - E[X]^2$$

$$= E_y[E_x[x^2|y]] - E[X]^2 = E_y[E_x[x^2|y] - E_x[x|Y]^2 + E_X[X|Y]^2] - E_Y[E_X[X|Y]]^2$$

$$= E_Y[E_X[x^2|Y] - E_x[X|Y]^2] + E_y[E[X|Y]^2] - E_Y[E_X[X|Y]]^2$$

$$= E_Y[var_X[X|Y]] + Var_y[E[X|Y]]$$

2 Maximum likelihood

(1)

$$\arg\max_{\theta} l(\theta) = \arg\max_{\theta} \sum_{i=1}^{n} \log f(x_i;\theta)$$

$$= \arg\max_{\theta} -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\alpha)^2$$

$$0 = \frac{\partial l(\theta)}{\partial \alpha} = \sum_{i=1}^{n}(x_i-\alpha) \Rightarrow \alpha = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \#$$

$$0 = \frac{\partial l(\theta)}{\partial \sigma^2} = \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\left(\sum (x_i-\alpha)^2\right) \Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^2,$$

if we rule out $\sigma^2 = 0$

(2)

$$l(\theta) = \sum_{i=1}^{n} \ln f(x_i;\mu,\Sigma) = -\frac{nd}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-u)^T\Sigma^{-1}(x_i-u)$$

$$\Rightarrow 0 = \frac{\partial l(\theta)}{\partial u} = \Sigma^{-1}(x-u) \Rightarrow u^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \#$$

3 Weighted Least Squares Regression

(1)

Let $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, $W = \begin{bmatrix} b \\ B^T \end{bmatrix}$, $C = \begin{bmatrix} c_1 & & 0 \\ 0 & c_2 & \vdots \\ 0 & \cdots & c_n \end{bmatrix}$

$\Rightarrow$ objective function become $\arg\min_{W} \|C(Y-XW)\|_2^2$

$\Rightarrow$ by taking derivative w.r.t. W, then we have

$$0 = C^TX^TC(Y-XW) \Rightarrow W = (C^TX^TCX)^{-1}C^TX^TCY \#$$

when $c_i = 1 \ \forall i \in 1 \sim n$

then solution become $W = (X^TX)^{-1}X^TY$

Which is equivalent to the solution of $\arg\min_{W} \|Y - XW\|_2^2$
then if X is a R.V. who is i.id on $x_1, x_2, \cdots x_n$
based on the Maximum likelihood method, where $W^* \in \theta$
$\arg\max_{\theta} p(\theta|X)$ if we don't have posi knowledge of $p(\theta)$.

$$= \arg\max \log p(X|\theta)$$

$$= \log\left(\prod_{i=1}^{n} p(y_i | x_i, \theta)\, p(x_i|\theta)\right)$$

$$= \log\left[\prod_{i=1}^{n} p(Wx_i + \varepsilon)\, p(x_i|\theta)\right]$$, since x_i is not depend on θ

$$\propto \log\left(\prod_{i=1}^{n} p(\varepsilon | x_i, \theta)\right).$$

$$\propto \log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - Wx_i)^2}{2\sigma^2}\right)\right) \propto -\sum_{i=1}^{n} \frac{1}{2\sigma^2}(y_i - Wx_i)^2$$

$$\propto -\sum_{i=1}^{n} (y_i - Wx_i)^2$$

$$\Rightarrow \arg\max_{\theta} \log p(x|\theta) = \arg\min_{W} \|y_i - Wx_i\|_2^2$$

(2)

if each ε_i has different variance
then

$$\arg\max_{\theta} \log p(X|\theta) = \arg\min_{W} \left\| \frac{c_i}{\sigma_i}(y_i - Wx_i) \right\|_2^2$$

$\Rightarrow$ the same form as equation (6) on problem 3 (1)

As a result, with the same variance, the answer will be general linear regression, but with different variance the answer will be local weighted linear regression.

$\Rightarrow$ Let $\hat{C} = \begin{bmatrix} \frac{c_1}{\sigma_1} & & 0 \\ & \ddots & \\ 0 & & \frac{c_n}{\sigma_n} \end{bmatrix}$ by problem 3. (1)

$$W = (\hat{C}^T X^T \hat{C} X)^{-1} \hat{C}^T X^T \hat{C} Y$$

4. Bias-Variance Decomposition

(1)

$$E_{Y|X,D}[(y-f_D(x))^2] = E_{Y|X,D}[(y-f(x)+f(x)-f_D(x))^2]$$

$$= E_{Y|X}[(y-f(x))^2] + E_D[(f(x)-f_D(x))^2] + 2E[(y-f(x))(f(x)-f_D(x))]$$

(where $E_{Y|X,D}[(y-f(x))(f(x)-f_D(x))] = E_{y|x}[y]f(x) - f(x)^2 - E_{y|x,D}[y f_D(x)] + f(x) E_D[f_D(x)]$

since $y = f(x) + \varepsilon$, & ε is zero mean.

$$= f(x)^2 - f(x)^2 - f(x) E_D[f_D(x)] + f(x) E_D[f_D(x)] = 0$$)

$$= E_{Y|X}[(y-f(x))^2] + E_D[(f(x)-f_D(x))^2]$$ *

(2)

$$E_{Y|X}[(y-f(x))^2] = E_\varepsilon[(f(x)+\varepsilon-f(x))^2] = E[\varepsilon^2]$$

also $\because E[\varepsilon]=0$

$$\therefore = E[\varepsilon^2] - E[\varepsilon]^2 = var[\varepsilon]$$ #

(3)

$$E_D[(f(x)-f_D(x))^2] = E_D[(f(x)-E_D[f_D(x)] + E_D[f_D(x)] - f_D(x))^2]$$

$$= (f(x)-E_D[f_D(x)])^2 + E_D[(E_D[f_D(x)] - f_D(x))^2] + 2\{\cancel{f(x)E_D[f_D(x)] - f(x)E_D[f_D(x)] - E_D[f_D(x)]^2 + E_D[f_D(x)]^2}\}$$

$$= (f(x)-E_D[f_D(x)])^2 + E_D[(E_D[f_D(x)] - f_D(x))^2]$$

(4)

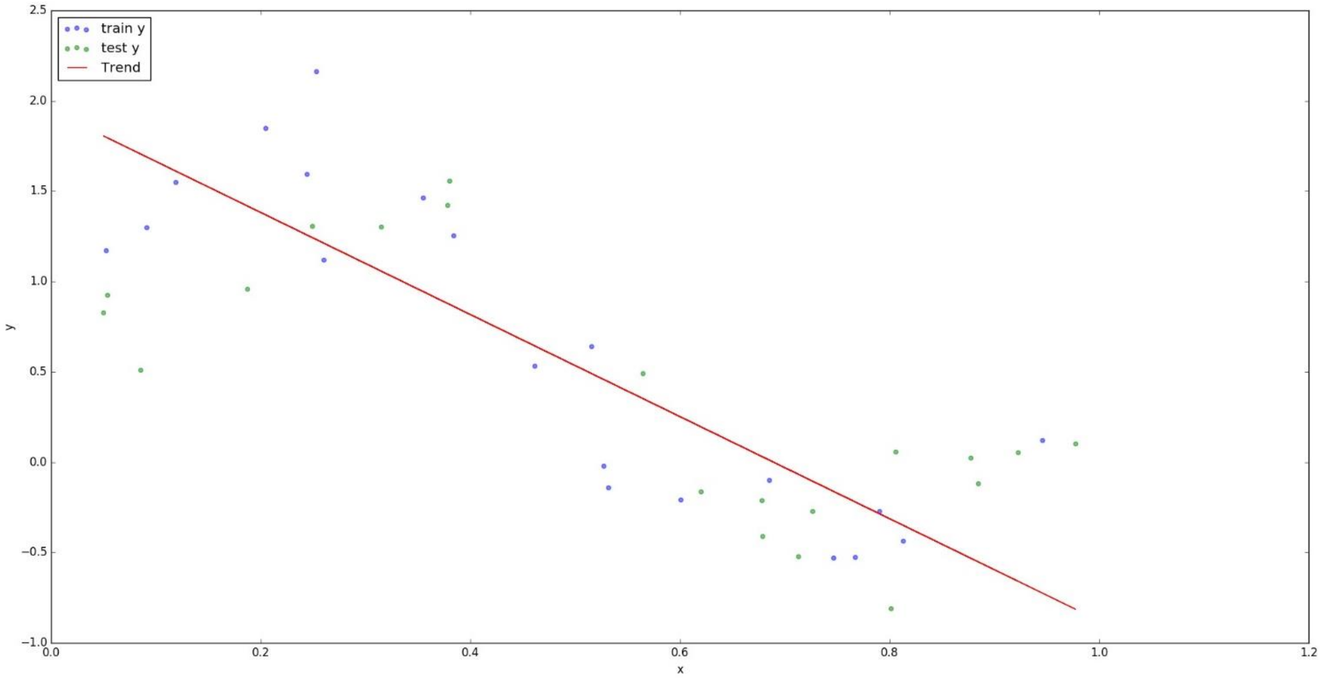
$$E_{X,Y,D}[(y-f_D(x))^2] \overset{by (1)(2)}{=} var[\varepsilon] + E_D[(f(x)-f_D(x))^2]$$

$$\overset{by (3)}{=} var[\varepsilon] + bias(x) + variance(x)$$ #

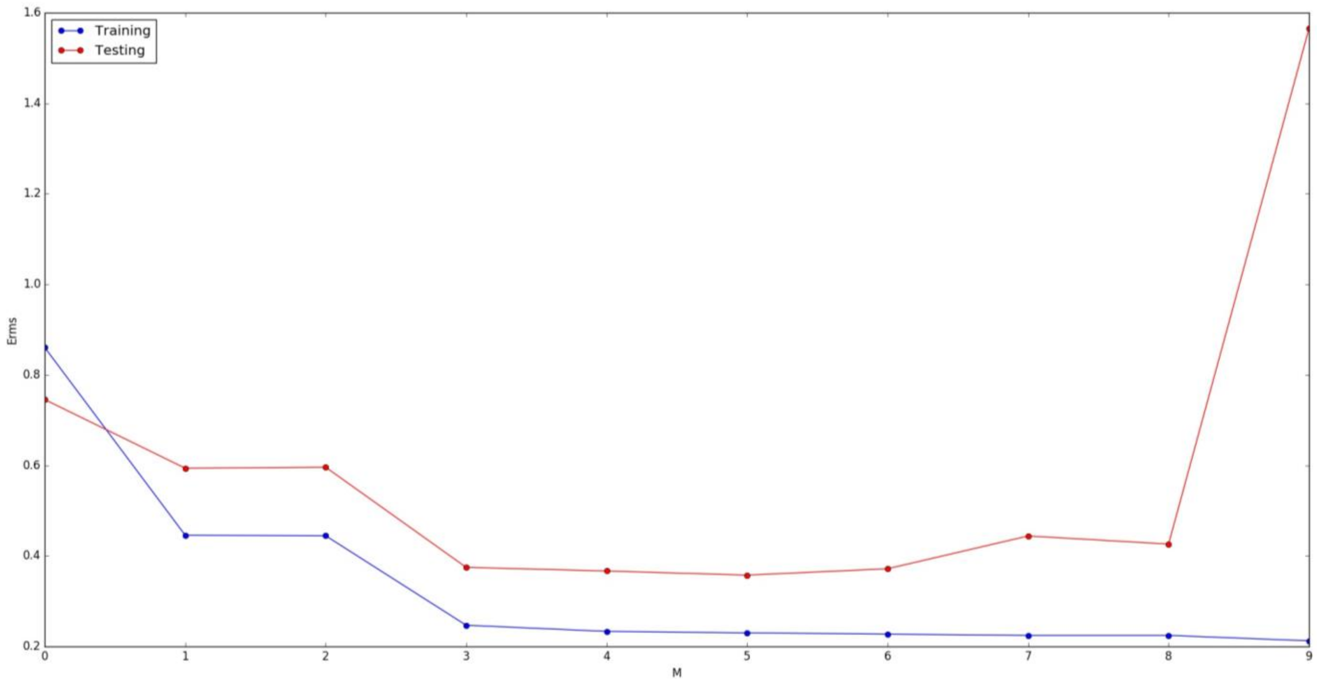
P5 Linear Regression

(1)

$$w = \begin{bmatrix} 1.9468 \\ -2.8241 \end{bmatrix}$$



(2)

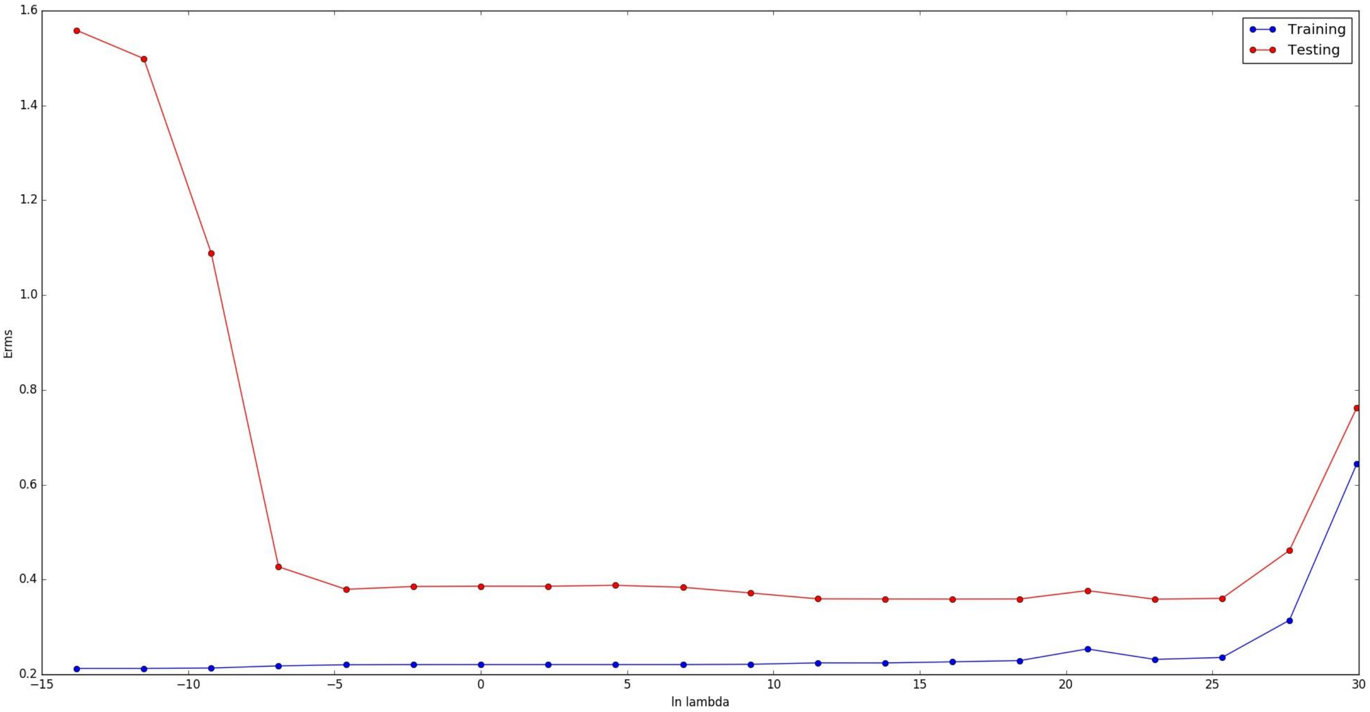


(3)

In this question, I set $\lambda = 10^{-6}: 10^{13}$. The close form for regularized linear regression is

$$w = y\Phi^T(\Phi\Phi^T + \lambda I)^{-1}$$

Where Φ is a 9$^{\text{th}}$ degree polynomial of x.



Training
Testing
Erms
ln lambda
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
−15
−10
−5
0
5
10
15
20
25
30