1., 4., 5.

MACHINE LEARNING ASSIGNMENT.

1. t_n = true label

x_n = feature vector

$r_n > 0 \rightarrow$ importance of n^{th} example.

$$E(w) = \frac{1}{2} \sum_{n=1}^{N} r_n \left(t_n - w^T \phi(x_n)\right)^2 .$$

Taking gradient,

$$\sum_{n=1}^{N} r_n \left\{ t_n - w^T \phi(x_n) \right\} \phi(x_n)^T = 0$$

$$\sum_{n=1}^{N} r_n t_n \phi(x_n)^T = \sum_{n=1}^{N} r_n w^T \phi(x) \phi(x_n)^T$$

If
$$\phi = \begin{pmatrix} \phi_0(x_1) & \phi_1(x_1) & \cdots & \phi_{M-1}(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \cdots & \phi_{M-1}(x_2) \\ \vdots & & & \\ \phi_0(x_N) & \phi_1(x_N) & \cdots & \phi_{M-1}(x_N) \end{pmatrix}$$

Taking
$$\phi_r = \begin{pmatrix} \sqrt{r_1}\,\phi_0(x_1) & \sqrt{r_1}\,\phi_1(x_1) & \cdots & \sqrt{r_1}\,\phi_{M-1}(x_1) \\ \sqrt{r_2}\,\phi_0(x_2) & \sqrt{r_2}\,\phi_1(x_2) & \cdots & \sqrt{r_2}\,\phi_{M-1}(x_2) \\ \vdots & & & \\ \sqrt{r_N}\,\phi_0(x_N) & \sqrt{r_n}\,\phi_1(x_N) & \cdots & \sqrt{r_N}\,\phi_{M-1}(x_N) \end{pmatrix}$$

$$w_{ML} = (\phi_r^T \phi_r)^{-1} \phi^T (t \odot r)$$

$\odot \rightarrow$ element wise product.

4. $P(w|t, x, \alpha) \rightarrow$ predictive distribution of weight vector w, for MAP inference of linear regression.

For a given value of x, the corresponding value of t has a gaussian distribution with mean value $y(x, w)$

$$p(t \mid x, w, \beta) = \mathcal{N}(t \mid y(x,w), \beta^{-1})$$

where $\beta^{-1} = \sigma^2$ $y(x,w) = \mu_1$

How

that

$$p(\bar{t} \mid \bar{x}, w, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n \mid y(x_n, w), \beta^{-1}).$$

$$\ln p(\bar{t} \mid \bar{x}, w, \beta) = -\frac{\beta}{2} \underbrace{\sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2}_{①} + \frac{N}{2} \ln \beta - \frac{N}{2} \ln 2\pi.$$

minimizing over β.

$$\frac{1}{\beta_{ML}} = \frac{1}{N} \sum_{n=1}^{N} \{y(x_n, w_{ML}) - t_n\}^2$$

$w_{ML} \rightarrow$ minimization of ①. (least square errors).

using predictive distributions,

$$p(t \mid x, w_{ML}, \beta_{ML}) = \mathcal{N}(t \mid y(x, w_{ML}), \beta^{-1}_{ML})$$

prior distribution of w w.r.t α,

$$p(\bar{w} \mid \alpha) = \mathcal{N}(\bar{w}, 0, \alpha^{-1} I)$$

$$= \left(\frac{\alpha}{2\pi}\right)^{\frac{M+1}{2}} e^{\left(-\frac{\alpha}{2} w^T w\right)}$$

From probability, $p(x,y) = p(x \mid y) p(y)$

$p(x \mid y) = p(x)$

for independent (x,y)

$$\therefore p(w \mid \bar{x}, \bar{t}, \alpha, \beta) \cancel{\propto} \propto p(\bar{t} \mid \bar{x}, w, \beta)\, p(w \mid \alpha)$$

$w \rightarrow$ independent of β and x.

$$p(\bar{t} \mid \bar{x}, w, \beta) \rightarrow \mathcal{N}(\mu_1, \sigma_1^2)$$

$$p(w \mid \alpha) \rightarrow \mathcal{N}(\mu_2, \sigma_2^2)$$

$$\Rightarrow p(w \mid \bar{x}, \bar{t}, \alpha, \beta) = \mathcal{N}(\underset{\downarrow}{\cancel{\mu_1 + \mu_2}}, \underset{\downarrow}{\cancel{\sigma_1^2 + \sigma_2^2}}).$$

$$\frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_2^2 + \sigma_1^2} \qquad \sqrt{\frac{\sigma_2^2 \sigma_1^2}{\sigma_2^2 + \sigma_1^2}}$$

5. M variables, D^{th} degree polynomial.

A Homogenous polynomial is one of the form such that $(x_1 \dots x_m) \to$ variables of degree 'd'

$$P(\lambda x_1 \dots \lambda x_n) = \lambda^d P(x_1 \dots x_n) -$$

$$P(x_1 \dots x_n) = 0 \Rightarrow P(\lambda x_1 \dots \lambda x_n) = 0$$

given a polynomial ring $R = K[x_1 \dots x_n]$ over a field K

the homogenous polynomials of degree d form a vector space Rd

R is the direct sum of Rd (non negative integers d)

(all polynomials can be represented as a sum of homogenous polynomials).

Dim (vector space Rd) $\to$ no of different monomials of degree d in M variables.

$\to$ which is the maximal number of non zero terms in a homogenous polynomials of degree d in M variabl

from binomial theorem for degree d, no of terms of homogenous poly $\to \binom{d+n-1}{n-1} = \frac{(d+M-1)!}{d!\,(M-1)!}$

for the polynomial of degree D. no of terms

$$= \sum_{d=1}^{D} \frac{(M+d-1)!}{(M-1)!\,d!}$$

2. Training set error with a random division of 60:40

S.NO	Lambda value	M.S. Error(E)
1.	0.00001	6.3969
2.	0.0001	6.3969
3.	0.001	6.3969
4.	0.01	6.3969
5.	0.1	6.3969
6.	1	6.3981
7.	10	6.4047
8	100	6.4391
9	1000	6.5201

Test set error:

S.NO	Lambda value	M.S. Error(E)
1.	0.00001	5.2902
2.	0.0001	5.2902
3.	0.001	5.2902
4.	0.01	5.2902
5.	0.1	5.2896
6.	1	5.2622
7.	10	
8	100	5.1991
9	1000	5.2024

3. **Weka results** with a random division of 60:40

Mean absolute error 4.4703

Root mean squared error 6.0108

Relative absolute error 68.5806 %

Root relative squared error 66.9563 %

7. Mean squared errors:

	Lambda1	Lambda2	Lambda3	Lambda4	Lambda5	Lambda6	Lambda7
a1	0.106766237267103	0.0944801227477411	0.109713867832934	0.100697142166924	0.198879695313357	0.320900675405592	0.323655044354624
a2	0.509126982108216	0.486898005162651	0.426046683884917	0.446444681619496	0.518001051186376	0.527320309710945	0.642683348350416
a3	0.952905300508052	1.01499511263051	0.888073616186469	1.01513912269988	0.827790819875525	1.04173110348522	0.980479018659859
a4	1.63278214089663	1.62122499438806	2.00100738565873	1.95366832890328	1.63424671219276	2.06143587584463	1.99208300491247
a5	8.40552901392863	9.18652425083571	10.7510405697484	7.81720856478126	9.44903406962857	10.0252928121391	9.13949004483804

8.

Bias of the Mean =

0.0155

Bias of the Variance =

-0.2621

Variance of the Mean =

0.2152

Variance of the Variance =

0.7529