Tutorial 8 : Linear Regression

1 Theory

Notation

Y - Output space : $y \in Y$ (Usually $Y \subseteq \mathbb{R}$).

X - Input space : $x \in X$ (Usually $X = \mathbb{R}^d$).

D - $D = \{x_k, y_k\}_{k=1}^n$ Training set, pairs of inputs and outputs (labels).

Statistical model

We assume that the following statistical model holds:

$$y = f_0(x) + \epsilon$$

where $f_0(x)$ is an unknown, deterministic function, and ϵ is a random variable representing an error, independent of x and with $E[\epsilon] = 0$. Therefore,

$$E[y|x] = f_0(x) + E[\epsilon] = f_0(x).$$

Goal

Learn a function $\hat{f}(x)$ which is the best possible approximation to $f_0(x)$. $\hat{f}(x)$ is called the **regression function**.

Linear regression function

We try to find the best approximation to $f_0(x)$ by a linear combination:

$$\hat{f}_w(x) = w^T \phi(x) = \sum_{m=1}^{M} w_m \phi_m(x)$$

where $\{\phi_m(x)\}_{m=1}^M$ is a pre-determined set of basis functions, and $w = (w_1, w_2, ..., w_M)^T \in \mathbb{R}^M$ is the vector of parameters we need to learn. We learn the parameters according to the squared error criterion, using the training set:

$$w^* = \underset{w \in \mathbb{R}^M}{\arg\min} \sum_{k=1}^{n} \left(y_k - \hat{f}_w(x)\right)^2 = \sum_{k=1}^{n} \left(y_k - w^T \phi(x)\right)^2.$$

2 Practice

Question 1

(a) Show that the optimal solution is given by $w = Q^{-1}b$ where

$$b = \sum_{k=1}^{n} \phi(x_k)y_k, \ Q = \sum_{k=1}^{n} \phi(x_k)\phi(x_k)^T,$$

and $\phi(x) = \big(\phi_1(x), \phi_2(x), ..., \phi_M(x)\big)$, as long as Q is not singular. What happens when Q is singular?

(b) Show that it is possible to express Q and b in the following manner:

$$b = H^T y, \ Q = H^T H,$$

where

$$y \triangleq (y_1, y_2, ..., y_n)^T, \ H \triangleq \big(\phi(x_1), \phi(x_2), ..., \phi(x_n)\big)^T \in \mathbb{R}^{n\times M}$$

(c) Write down H explicitly for a linear model, that is for $\phi_0(x) = 1, \ \phi_m(x) = x_m \ \ m = 1, ..., d$.

Solution

(a) Notice that

$$\begin{aligned}
\sum_{k=1}^{n} \Big(y_k - \hat{f}_w(x)\Big)^2 &= \sum_{k=1}^{n} y_k^2 - 2w^T\phi(x_k)y_k + \big(w^T\phi(k)\big)^2 \\
&= \sum_{k=1}^{n} y_k^2 - 2w^T\phi(x_k)y_k + \big(w^T\phi(k)\big)\big(w^T\phi(k)\big) \\
&= \sum_{k=1}^{n} y_k^2 - 2w^T\phi(x_k)y_k + w^T\phi(k)\phi(k)^T w \\
&= w^T Q w - 2w^T b + c
\end{aligned}$$

where

$$Q = \sum_{k=1}^{n} \phi(x_k)\phi(x_k)^T, \ b = \sum_{k=1}^{n} \phi(x_k)y_k, \ c = \sum_{k=1}^{n} y_k^2.$$

In order to find the minimal point, we set the derivative with respect to w to zero:

$$2Qw - 2b = 0 \ \Rightarrow \ Qw = b$$

where we use the fact that Q is a symmetric matrix. Assuming that Q is invertible we get that

$$w^* = Q^{-1}b.$$

Such a solution exists if Q is not singular, i.e, $\{\phi(x_k)\}_{k=1}^{n}$ spans $\mathbb{R}^M$. Generally, if Q is singular, the equation $Qw = b$ has either no solution, or infinite number of solutions. However, in our case, since b is a linear combination of the basis functions, there will always be infinite number of solutions when Q is singular. There are a few techniques for choosing one solution out of the infinite number. A common way is to use the generalized inverse, aka Moore-Penrose inverse or Pseudo-inverse. In this case the solution has a minimal norm $||w||_2^2$.

(b) Note that

$$b = \sum_{k=1}^{n} \phi(x_k) y_k = \begin{bmatrix} \phi(x_1) \ \phi(x_2) \ \cdots \ \phi(x_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = H^T y.$$

$$Q = \sum_{k=1}^{n} \phi(x_k) \phi(x_k)^T = \begin{bmatrix} \phi(x_1) \ \phi(x_2) \ \cdots \ \phi(x_n) \end{bmatrix} \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix} = H^T H.$$

Therefore, we get

$$w = Q^{-1} b = (H^T H)^{-1} H^T y \triangleq H^{\dagger} y$$

Note that $H^{\dagger} = (H^T H)^{-1} H^T$ is the pseudo-inverse of H.

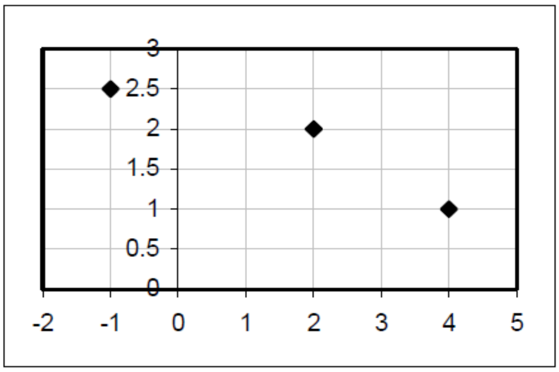
(c) For the linear model, H is

$$\begin{bmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nd} \end{bmatrix}$$

Question 2

Let $X = \mathbb{R}$ and $Y = \mathbb{R}$ and consider the following training set

$$D = \{(-1, 2.5),\ (2, 2),\ (4, 1)\}.$$



(a) Write down a liner model for regression for this case. Find the optimal parameter vector for this model.

(b) Write down a second order polynomial model for regression for this case. Find the optimal parameter vector for this model.

(c) Write down a third order polynomial model for regression for this case. Is there a single solution for the parameter vector? Find the parameter vector for the two following cases:

(i) $w_0 = 0$.

(ii) $w_2 = 0$.

Solution

(a) In this case,

$$y = \begin{bmatrix} 2.5 \\ 2 \\ 1 \end{bmatrix}, \; H = \begin{bmatrix} 1 & -1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}.$$

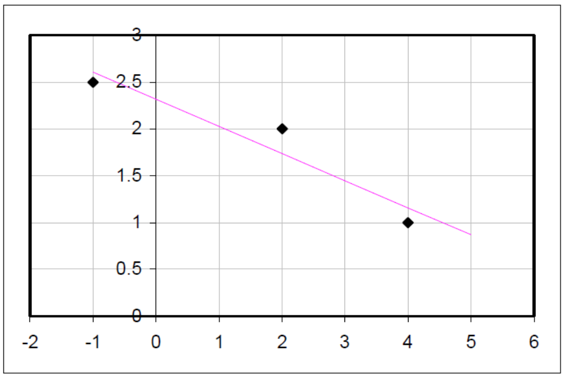
Therefore,

$$H^T H = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 5 \\ 5 & 21 \end{bmatrix}$$

$$\Rightarrow \quad (H^T H)^{-1} H^T = \begin{bmatrix} 0.6842 & 0.2895 & 0.0263 \\ -0.2105 & 0.0263 & 0.1842 \end{bmatrix}$$

$$\Rightarrow \quad w^* = \begin{bmatrix} 0.6842 & 0.2895 & 0.0263 \\ -0.2105 & 0.0263 & 0.1842 \end{bmatrix} \begin{bmatrix} 2.5 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2.3158 \\ -0.2895 \end{bmatrix}$$

$$\Rightarrow \quad f(x) = 2.3158 - 0.2895x.$$

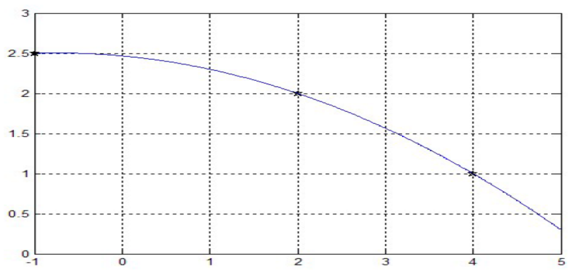


(b) Second order polynomial model $\hat{f}_w(x) = w_0 + w_1 x + w_2 x^2$ that is $\phi(x) = (1, x, x^2)^T$. Hence, we have

$$y = \begin{bmatrix} 2.5 \\ 2 \\ 1 \end{bmatrix}, \; H = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix}$$

$$\Rightarrow \quad w^* = \begin{bmatrix} 2.4667 \\ -0.1 \\ -0.0667 \end{bmatrix}$$

$$\Rightarrow \quad \hat{f}_w(x) = 2.4667 - 0.1x - 0.0667x^2$$



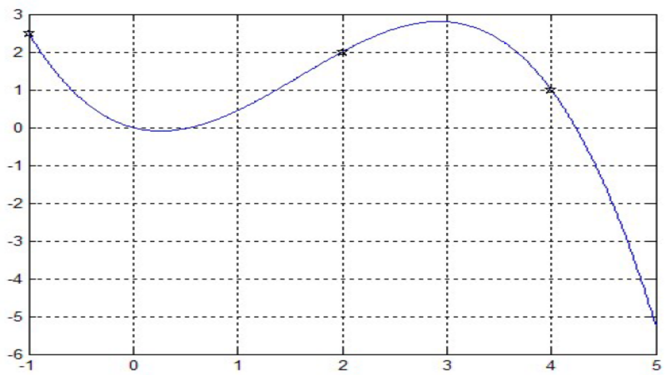
(c) Third order polynomial model $\hat{f}_w(x) = w_0 + w_1 x + w_2 x^2 + w_3 x^3$, that is $\phi(x) = (1, x, x^2, x^3)^T$. Thus, we get

$$H = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 2 & 4 & 8 \\ 1 & 4 & 16 & 64 \end{bmatrix}.$$

It can be verified that the rank of H is 3, meaning that H is singular. This makes sense since we have more variables then samples, so there is an infinite number of solutions.

(i) $w_0 = 0$, hence, $\hat{f}_w(x) = w_1 x + w_2 x^2 + w_3 x^3$ and we get $\phi(x) = (x, x^2, x^3)^T$.

$$w^* = \begin{bmatrix} 0 \\ -0.7167 \\ 1.4750 \\ -0.3083 \end{bmatrix} \Rightarrow \hat{f}_w(x) = -0.7167x + 1.4750x^2 - 0.3083x^3.$$



(ii) $w_2 = 0$, hence, $\hat{f}_w(x) = w_0 + w_1 x + w_3 x^3$ and we get $\phi(x) = (1, x, x^3)^T$.

$$w^* = \begin{bmatrix} 2.36 \\ -0.1267 \\ 0 \\ -0.0133 \end{bmatrix} \Rightarrow \hat{f}_w(x) = 2.36 - 0.1267x - 0.0133x^3.$$