Machine Learning

V. Adamchik CSCI 567 Spring 2019

Discussion 3 University of Southern California

Linear Regression

Linear Model

Given some points in D-space, $x \in \mathbb{R}^D$

The goal is to fit them with a hyperplane

$$f(x) = w_0 + w_1 x_1 + \ldots + w_D x_D$$

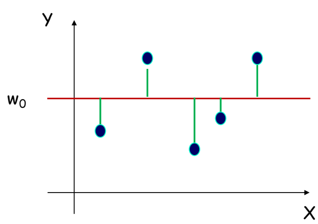
In the lecture you considered a particular case of D = 0 using L_2 norm.

And you proved that w_0 is the average wrt y.

Problem 1

Minimize the total absolute error (L_1 norm) of linear regression when D = 0:

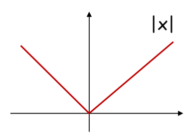
$$\min_{w_0} \sum_n |w_0 - y_n|$$



Solution

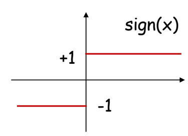
We can proceed as in lecture by finding a derivative

What is $\frac{d}{dx}|x|$



Applying derivative, we get

$$\frac{d}{dx}\sum_n |w_0 - y_n| = \sum_n sign(w_0 - y_n) = 0$$



This is zero if the number of + is equal to the number of -. Let us assume that N is odd.

It happens when w_0 is the median.

Problem 2

Suppose A is a square matrix. Then A is singular if and only if $\lambda = 0$ is an eigenvalue of A. Prove this statement.

Solution.

λ is an eigenvalue of A if $Ax = \lambda x$ for $x \neq 0$

Since A is singular it means that

$$A x = 0 \text{ for } x \neq 0.$$

This can be written as $Ax = 0x$.

It follows $\lambda=0$ is an eigenvalue.

Problem 3

Suppose A is a square nonsingular matrix and λ is an eigenvalue of A. Then $1/\lambda$ is an eigenvalue of the matrix A^{-1}.

Solution.

λ is an eigenvalue of A if $Ax = \lambda x$.

Note A^{-1} exists since A is nonsingular.

Consider $A^{-1} x = A^{-1} (1/\lambda \, \lambda x) = 1/\lambda \, A^{-1} (\lambda x) = 1/\lambda \, (A^{-1} A) x$

We proved $A^{-1} x = 1/\lambda \, x$.

Problem 4

Find eigenvalues and eigenvectors for

$$A = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$

Solution.

Rewrite $Ax = \lambda x$ as $(A - \lambda I) x = 0$

$$A - \lambda I = \begin{pmatrix} 2-\lambda & 3 \\ 2 & 1-\lambda \end{pmatrix}$$

Compute its determinate $|A - \lambda I| = \lambda^2 - 3\lambda - 4 = (\lambda - 4)(\lambda + 1)$

We found two eigenvalues: $\lambda = 4$ and $\lambda = -1$.

Next, find an eigenvector for $\lambda = 4$.

Solution

Let $\lambda = 4$.

$$A - \lambda I = \begin{pmatrix} 2-\lambda & 3 \\ 2 & 1-\lambda \end{pmatrix} = \begin{pmatrix} -2 & 3 \\ 2 & -3 \end{pmatrix}$$

$$\begin{pmatrix} -2 & 3 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

It follows, $2 v_1 = 3 v_2$.

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = k \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

where k is any number.

Solution

Let λ = -1.

$$A - \lambda I = \begin{pmatrix} 2-\lambda & 3 \\ 2 & 1-\lambda \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 2 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 3 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

It follows, $v_1 = - v_2$.

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = k \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

where k is any number.

For most applications we normalize the eigenvectors $x^T x = 1$.

The Eigen-Decomposition

Traditionally, we put together eigenvectors in matrix denoted U. Each column of U is an eigenvector of A.

Using eigenvectors U and eigenvalues Λ (a diagonal matrix) we rewrite $Av = \lambda v$ as

$$AU = U\Lambda \quad \text{or} \quad A = U\Lambda U^{-1}$$

For the previous example:

$$A = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -2 & 3 \end{pmatrix} \frac{1}{5}$$

Special Case

In lecture the matrix $A = X^TX$ is a real symmetric matrix

$$X^TX = (X^TX)^T$$

If A is symmetric then its eigenvectors are orthogonal to each other (problem 5). Thus,

$$UU^T = U^TU = I$$

It follows, $\boxed{A = U\Lambda U^{-1} = U\Lambda U^T = U^T\Lambda U}$

Example,

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \frac{1}{2}$$

Problem 5

The eigenvectors of a symmetric matrix A corresponding to different eigenvalues are orthogonal to each other. Prove this statement.

Solution.

$$A x_1 = \lambda_1 x_1 \qquad x_2^T A x_1 = \lambda_1 x_2^T x_1$$

$$A x_2 = \lambda_2 x_2 \qquad x_1^T A x_2 = \lambda_2 x_1^T x_2$$

Subtracting from one another:

$$0 = x_2^T A x_1 - x_1^T A x_2 = (\lambda_1 - \lambda_2) x_1 x_2$$

Each term on the left hand side is a scalar and since A is symmetric $x_2^T A x_1 = (x_1^T A x_2)^T$, the left hand side is equal to zero.

Problem 6

Let u, x∈$\mathbb{R}^n$ are column vectors, A is (nxn) matrix.

Task 1. Compute $\frac{\partial}{\partial x} u^{\mathrm{T}} x$ and $\frac{\partial}{\partial x} x^{\mathrm{T}} u$

Task 2. Compute $\frac{\partial}{\partial x} \|x\|_2^2$

Task 3. Compute $\frac{\partial}{\partial x} x^{\mathrm{T}} A\, x$

Solution (task 1)

$$\frac{\partial}{\partial x} u^{\mathrm{T}} x = \left(\frac{\partial}{\partial x_1} u^{\mathrm{T}} x, \ldots, \frac{\partial}{\partial x_{\mathrm{n}}} u^{\mathrm{T}} x\right)^{\mathrm{T}}$$

$$u^{\mathrm{T}} x = \sum_{i=1}^{n} u_{\mathrm{i}}\, x_{\mathrm{i}} \quad = x^{\mathrm{T}} u$$

$$\frac{\partial}{\partial x} u^{\mathrm{T}} x = (u_1, \ldots, u_{\mathrm{n}})^{\mathrm{T}} = u$$

$$\frac{\partial}{\partial x} x^{\mathrm{T}} u = u$$

Solution (task 2)

$$\|x\|_2^2 = x^{\mathrm{T}} x = \sum_{i=1}^{n} x_{\mathrm{i}}\, x_{\mathrm{i}} = \sum_{i=1}^{n} x_{\mathrm{i}}^2$$

$$\frac{\partial}{\partial x_k} x^{\mathrm{T}} x = \frac{\partial}{\partial x_k} \sum_{i=1}^{n} x_{\mathrm{i}}^2 = 2x_k$$

$$\frac{\partial}{\partial x} x^{\mathrm{T}} x = (2x_1, \ldots, 2x_n)^{\mathrm{T}} = 2x$$

Solution (task 3)

see task 1

Using the product rule

($\overline{x}$ treated a constant):

$$\frac{\partial}{\partial x} x^T A x = \frac{\partial}{\partial x} x^T (A\overline{x}) + \frac{\partial}{\partial x} (\overline{x}^T A) x =$$

$$= \frac{\partial}{\partial x} x^T (A\overline{x}) + \frac{\partial}{\partial x} (A^T \overline{x})^{\boldsymbol{T}} x =$$

$$= Ax + A^T x \quad = (A + A^T)\, x$$