LECTURE 29

I. The Least Squared Approximation:
Its cognitive value in data mining.

II. Examples of Least Squares Problems
(i) Algebraic solutions
(ii) Geometric solutions.

III. The Fundamental Theorem:

Orthogonal Projection = Optimal Projection.

I. THE LEAST SQUARES APPROXIMATION:

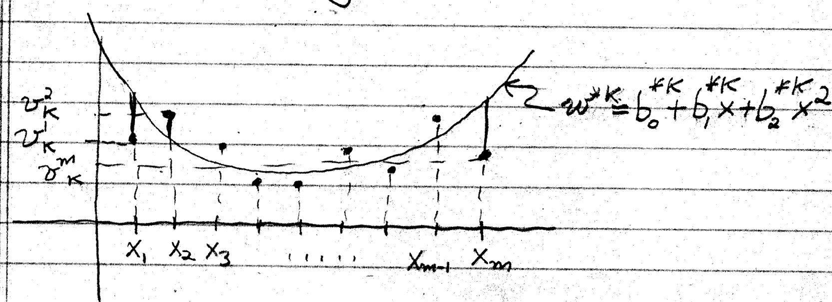
Of What Value Is it? Why?

Least squares approximations condense datasets into a cognitively managable, quantitative form whose value arises as a means to an end in data mining.

The cognitive elements of this condensation process vary from one experiment/observation to another, even though the experiments/observations are identically designed/executed (within the context of technological limitations).

These condensations [i.e. the to-be-calculated least squares approximations] have distinct but commensurable identities. In probability theory they are expressed quantitatively as "random" (or "stochastic") variables. In this theory one defines a probability function on this domain of random variables, and thereby arrives at potentially valuable knowledge about the aggregate of least squares approximations.

in the best way.



Here "best" means

$$\|\vec{v}_K - \vec{w}\|^2 \geq \|\vec{v}_K - \vec{w}^{*K}\|^2$$

for all

$$\vec{w} = \{ b_0 + b_1 x_i + b_2 (x_i)^2 \}_{i=1}^{m}$$

i.e for all

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \in R^3$$

a) Algebraically, the task is to find the "best" solution to the (inconsistent) K^{th} system (solve)

$$w(x_i) = v_K^i \qquad i = 1, 2, \ldots, m$$

SOLUTION: Rewrite this system in the form

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix}}_{\vec{u}_1\ \vec{u}_2\ \vec{u}_3 \in W_3} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \underbrace{\begin{bmatrix} v_K^1 \\ v_K^2 \\ \vdots \\ v_K^m \end{bmatrix}}_{\vec{v}_K \in V = R^m}$$

and then, by projecting $\vec{v}_K$ onto the subspace

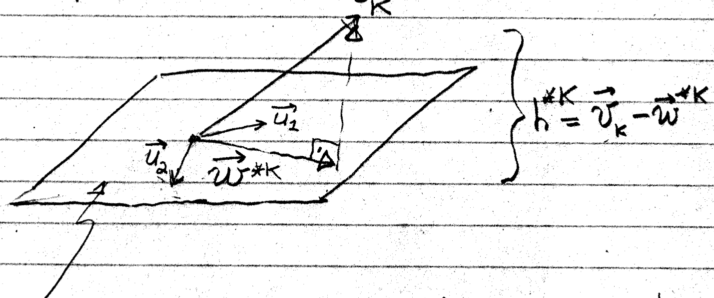
$$W_3 = Sp(\vec{u}_1, \vec{u}_2, \vec{u}_3) \subset V = R^3,$$

determine the optimal ("least squares") solution

$$\begin{bmatrix} b_0^{*K} \\ b_1^{*K} \\ b_2^{*K} \end{bmatrix}$$

corresponding to the K^{th} given data set $(\vec{x}, \vec{v}_K)$.

$V = R^m$



$W_N = Sp(\vec{u}_1, \ldots, \vec{u}_N)$ (N=3 in the preceding example)

b) <u>Geometrically</u>, the task is to find the <u>best vector</u>

$$\vec{w}^{*K} = \underbrace{b_0^{*K}\vec{u}_1}_{\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}} + \underbrace{b_1^{*K}\vec{u}_2}_{\begin{bmatrix}x_1\\x_2\\\vdots\\x_m\end{bmatrix}} + \underbrace{b_2^{*K}\vec{u}_3}_{\begin{bmatrix}x_1^2\\x_2^2\\\vdots\\x_m^2\end{bmatrix}}$$

<u>Comment 1</u>: Instead of basing the optimation on a 3-parmeter polynomial, an alternative least squares is possible namely use 3-term combinations of some lin. indep. set of 3 functions.

In that case one would have

$$\vec{u}_1 = \begin{bmatrix} f_1(x_1) \\ f_1(x_2) \\ \vdots \\ f_1(x_m) \end{bmatrix} \quad \vec{u}_2 = \begin{bmatrix} f_2(x_1) \\ f_2(x_2) \\ \vdots \\ f_2(x_m) \end{bmatrix} \quad \vec{u}_3 = \begin{bmatrix} f_3(x_1) \\ f_3(x_2) \\ \vdots \\ f_3(x_m) \end{bmatrix}.$$

<u>Comment 2</u>:

Familiarity with the properties of the shortest line between a point and a plane suggests that the "error vector"

$$\vec{v}^K - \vec{w}^{*K} \equiv \vec{h}^{*K}$$

be perpendicular to every vector in the subspace W_N, i.e. $\vec{h}^{*K} \perp W_N$. This observation is made precise by the following fundamental

Fundamental Theorem 29.1 (Perpendicular distance = Minimum distance)

Given:

V is an inner product space

$W_N \subset V$ is a subspace of V.

$\vec{v}^K \in V$ any vector in V.

Conclusion:

$\exists$ a unique $w^{*K} \in W_N$ such that

$$\langle \underbrace{\vec{v}^K - \vec{w}^{*K}}_{\vec{h}^{*K}} | \vec{w} \rangle = 0 \quad \forall \vec{w} \in W_N$$

if and only if ($\Leftrightarrow$)

$\exists$ a unique $\vec{w}^{*K} \in W_N$ such that

$$\|\vec{v}^K - \vec{w}^{K*}\| \leq \|\vec{v}^K - \vec{w}\| \qquad \forall \vec{w} \in W_N$$

or equivalently

$\mathcal{E}_K^2(\vec{w}) \equiv \|v^K - w\|^2$ has a unique minimum at $\vec{w} = \vec{w}^{*K}$.

Proof: ($\Rightarrow$)

Let $\vec{w} \in W_N$. Then

$$\|\vec{v}^K - \vec{w}\| = \langle \vec{v}^K - \vec{w}^{*K} + \vec{w}^{*K} - \vec{w} | \vec{v}^K - \vec{w}^{*K} + \vec{w}^{*K} - \vec{w} \rangle$$

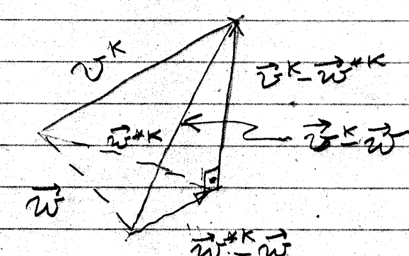
(★) $$= \|\vec{v}^K - \vec{w}^{*K}\|^2 + \underbrace{\langle \underbrace{\vec{w}^{*K} - \vec{w}}_{\in W_N} | \vec{v}^K - \vec{w}^{*K} \rangle}_{=0} + \underbrace{\langle \vec{v}^K - \vec{w}^{*K} | \underbrace{\vec{w}^{*K} - \vec{w}}_{\in W_N} \rangle}_{=0} + \|w^{*K} - w\|^2$$

given

Thus

$$\|v^K - \vec{w}^{*K}\| \leq \|\vec{v}^K - \vec{w}\| \qquad \forall \vec{w} \in W_N$$

Note: Equation (★) is a statement of the Pythagorean theorem



$$\boxed{\|\vec{v}^K - \vec{w}\|^2 = \|\vec{v}^K - \vec{w}^{*K}\|^2 + \|\vec{w}^{*K} - \vec{w}\|^2}$$

(hypotenuse)^2 =

Proof: ($\Leftarrow$) (in 4 steps)

Step I Introduce an o.N. basis for W_N

$$\vec{w} = c_1 e_1 + \cdots + c_N e_N, \text{ where } c_k = a_k + i b_k$$

and consider the error fn $\mathcal{E}^2(c_\ell)$
and for typographical reasons suppress the superscript K and the overarrow $\rightarrow$:

$$\mathcal{E}_K^2(c_\ell) = \|\vec{v}^K - \sum_{k=1}^N c_k e_k\|^2$$

$$= \langle v^K - c_k e_k | v^K - c_\ell e_\ell \rangle \equiv \mathcal{E}_K^2(a_\ell, b_\ell)$$

$$= \|v\|^2 + \bar{c}_k c_k - \langle v | e_\ell \rangle c_\ell - \bar{c}_k \langle e_k | v \rangle$$

$$= \|v\|^2 + \sum_{k=1}^N (a_k^2 + b_k^2) - \sum_{\ell=1}^N \langle v | e_\ell \rangle (a_\ell + i b_\ell) - \sum_{k=1}^N (a_k - i b_k) \langle e_k | v \rangle$$

Step II

$$\frac{\partial \mathcal{E}^2}{\partial a_k} = 2 a_k - \left(\langle v | e_k \rangle + \langle e_k | v \rangle \right) = 0$$

Hence: $a_k = \frac{1}{2} \langle v | e_k \rangle + \frac{1}{2} \overline{\langle v | e_k \rangle}$

$$\frac{\partial \mathcal{E}^2}{\partial b_k} = 2 b_k - i \left(\langle v | e_k \rangle - \langle e_k | v \rangle \right) = 0$$

Hence: $b_k = \frac{i}{2} \langle v | e_k \rangle - \frac{i}{2} \overline{\langle v | e_k \rangle}$

$$\left.\begin{aligned} &\text{Thus } a_k + i b_k \equiv c_k = \overline{\langle \vec{v}^K | \vec{e}_k \rangle} \\ &\vec{w}^{*K} = \sum_{k=1}^N \langle v^K | e_k \rangle \vec{e}_k \end{aligned}\right\} \underline{\text{unique vector in } W_N}$$

Step III Using this critical vector one finds

$$\langle \vec{v}^K - \vec{w}^{*K} | w \rangle = \langle \vec{v}^K - \sum c_k^{*K} \vec{e}_k | \sum c_\ell e_\ell \rangle$$

$$= \sum_\ell c_\ell \left(\langle \vec{v}^K | e_\ell \rangle - \sum_k \delta_{k\ell} \overline{c_k^{*K}} \right)$$

$$= \sum_\ell c_\ell \left(\langle \vec{v}^K | e_\ell \rangle - \langle \vec{v}^K | e_\ell \rangle \right)$$

$$= 0 \qquad \forall w \in W_N$$

Thus
$\vec{v}^K - \vec{w}^{*K} = \vec{h}^{*K}$ is $\perp$ W_N indeed

Step IV

$$\frac{\partial^2 \mathcal{E}^2}{\partial a_k^2} = 2 > 0, \quad \frac{\partial^2 \mathcal{E}^2}{\partial b_k^2} = 2 > 0.$$

Thus $\mathcal{E}_K^2$ is a minimum, i.e. $\vec{h}^{*K}$ is the shortest (= best) vector.

Comment
It is not necessary to use an o.N. basis for W_N in order to find the optimal.

This makes computations easy. This fact is expressed by the following theorem in LECTURE 30