EE306: Signals and Systems II

Lecture 10

MMSE linear estimator (Cont.)

* Now, suppose we have a vector of observations $\underline{Y} = [Y_1 \; Y_2 \; \cdots \; Y_n]^T$, and $\hat{X} = g(\underline{Y}) = \underline{a}^T \underline{Y} + b$.

→ Here, $e = E[(X-\hat{X})^2] = E[(X-\underline{a}^T\underline{Y}-b)^2]$

Thus, $b^* = E(X-\underline{a}^T\underline{Y}) = E(X) - \underline{a}^T E(\underline{Y}) = m_X - \underline{a}^T \underline{m}_Y$.

$\Rightarrow e = E[(X-m_X-\underline{a}^T(\underline{Y}-\underline{m}_Y))^2] \rightarrow \circledast$

→ Gradient and Hessian:

For $f(\underline{X}) = f(x_1, x_2, \ldots, x_n)$,

$$\nabla_{\underline{X}} f(\underline{X}) = \left[\frac{\partial f(\underline{X})}{\partial x_1} \; \frac{\partial f(\underline{X})}{\partial x_2} \; \cdots \; \frac{\partial f(\underline{X})}{\partial x_n}\right]^T \text{ and}$$

$$\nabla^2_{\underline{X}} f(\underline{X}) = \left[\nabla_{\underline{X}} \frac{\partial f(\underline{X})}{\partial x_1} \; \nabla_{\underline{X}} \frac{\partial f(\underline{X})}{\partial x_2} \; \cdots \; \nabla_{\underline{X}} \frac{\partial f(\underline{X})}{\partial x_n}\right]^T.$$

→ Back to $\circledast$: $\nabla_{\underline{a}} e$ is the grad. w.r.t. $\underline{a}$

$$\nabla_{\underline{a}} e = -2E\left[\left(X-m_X-\overbrace{\underline{a}^T(\underline{Y}-\underline{m}_Y)\right)(\underline{Y}-\underline{m}_Y)}^{(\underline{Y}-\underline{m}_Y)(\underline{Y}-\underline{m}_Y)^T\underline{a}}\right]$$

$$\nabla_{\underline{a}} e \Big|_{\underline{a}=\underline{a}^*} = \underline{0} \Rightarrow \underline{a}^* = K_{\underline{Y}}^{-1} E[(X-m_X)(\underline{Y}-\underline{m}_Y)]$$

Therefore, $g(\underline{Y}) = \left[K_{\underline{Y}}^{-1} E[(X-m_X)(\underline{Y}-\underline{m}_Y)]\right]^T (\underline{Y}-\underline{m}_Y) + m_X$.

→ observe orthogonality here too.

MMSE generalized estimator (also called non-linear)

* The goal is to find generic $g(Y)$ that minimizes the error, $\underset{g(\cdot)}{\operatorname{argmin}} E[(X-g(Y))^2]$

→ Using the total expect. theorem,

$$E_Y[E[(X-g(Y))^2|Y]] = E[(X-g(Y))^2] \Rightarrow$$

$$E[(X-g(Y))^2] = \int_y E[(X-g(Y))^2 | Y=y]\, f_Y(y)\, dy.$$

→ Given $Y=y$, $g(Y)$ becomes a const.

Thus, $g^*(Y) = \underline{E[X|Y=y]}$,

which is called the regression curve.

* Similarly, in the case of $g(\underline{Y})$ (vector of obs.'s) we have $g^*(\underline{Y}) = E[X|\underline{Y}=\underline{y}]$.

EX1 Let X and Y be joint RVs s.t.

$f_{X,Y}(x,y) = 2e^{-x}e^{-y}$, $0\leq y\leq x<\infty$, $f_{X,Y}(x,y)=0$, o.w.

(a) Find $f_X(x)$ and $f_Y(y)$.

(b) Find $E(X)$, $V(X)$, $E(Y)$, $V(Y)$, and $CV(X,Y)$.

(c) Optimal linear/generalized estim. for X given Y.

(d) Optimal linear/generalized estim. for Y given X.

Sol. (a) $f_X(x) = 2e^{-x}(1-e^{-x})$, $x\geq 0$, $f_Y(y) = 2e^{-2y}$, $y\geq 0$.

(b) $E(X)=\frac{3}{2}$, $V(X)=\frac{5}{4}$, $E(Y)=\frac{1}{2}$, $V(Y)=\frac{1}{4}$

$$CV(X,Y)=\frac{1}{4} \quad (\rho_{X,Y}=\frac{1}{\sqrt{5}}).$$

(c) Linear: $\hat{X}=\rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(Y-m_Y)+m_X=\frac{1}{\sqrt{5}}\frac{\sqrt{5}/2}{1/2}(Y-\frac{1}{2})+\frac{3}{2}$

$\Rightarrow \hat{X}=Y+1.$

Generalized: $\hat{x}=E[X|Y=y]=\int_{x=y}^{\infty}\frac{2e^{-x}e^{-y}}{2e^{-2y}}x\,dx$

$\Rightarrow \hat{x}=\int_{x=y}^{\infty}e^{y}xe^{-x}dx=y+1 \Rightarrow \hat{X}=Y+1.$ (linear)

(d) Linear: $\hat{Y}=\rho_{XY}\frac{\sigma_Y}{\sigma_X}(X-m_X)+m_Y=\frac{1}{\sqrt{5}}\frac{1/2}{\sqrt{5}/2}(X-\frac{3}{2})+\frac{1}{2}$

$\Rightarrow \hat{Y}=\frac{1}{5}(X+1).$

Generalized: $\hat{y}=E[Y|X=x]=\int_{0}^{x}\frac{2e^{-x}e^{-y}}{2e^{-x}(1-e^{-x})}y\,dy$

$\Rightarrow \hat{y}=\int_{0}^{x}\frac{ye^{-y}}{(1-e^{-x})}dy=1-\frac{xe^{-x}}{1-e^{-x}} \Rightarrow \hat{Y}=1-\frac{Xe^{-X}}{1-e^{-X}}.$ (non-linear)

The two estimators here differ remarkably.

Linear least squares (LLS)

* Suppose you have n points in a k-dimensional space. $n>k$

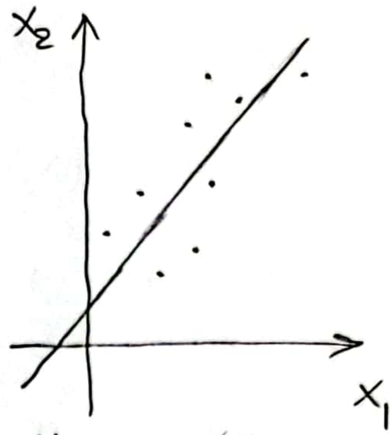
The points have the form:

$$\underline{x}_i^T=(x_{i,1},x_{i,2},\dots,x_{i,k}), \quad 1\le i\le n \rightarrow \text{row of } \underline{X}.$$

The LLS problem is the problem of finding a line that minimizes the error to these points. The line has the form: $\beta_1\underbrace{x_1}_{x_{i1}}+\beta_2\underbrace{x_2}_{x_{i2}}+\cdots+\beta_k\underbrace{x_k}_{x_{i,k}}=\underbrace{y}_{y_i}.$

This linear regression problem can be formulated as follows:

$$\underset{\underline{\beta}}{\arg\min} \|\underline{X}\,\underline{\beta}-\underline{y}\|_2.$$



Here $n=9$ and $k=2$.

Ex2 Ordinary least squares (OLS):

Let the matrix $\underline{X}$ have full rank:

$$\|\underline{X}\,\underline{\beta}-\underline{y}\|_2^2 = (\underline{X}\,\underline{\beta}-\underline{y})^T(\underline{X}\,\underline{\beta}-\underline{y})$$

$$= \underline{\beta}^T\underline{X}^T\underline{X}\,\underline{\beta} - \underline{\beta}^T\underline{X}^T\underline{y} - \underline{y}^T\underline{X}\,\underline{\beta} + \underline{y}^T\underline{y} \Rightarrow$$

$$\nabla_{\underline{\beta}}\|\underline{X}\,\underline{\beta}-\underline{y}\|_2^2 = 2\underline{X}^T\underline{X}\,\underline{\beta} - 2\underline{X}^T\underline{y} \Rightarrow$$

$$\nabla_{\underline{\beta}}\|\underline{X}\,\underline{\beta}-\underline{y}\|_2^2\Big|_{\underline{\beta}=\underline{\beta}^*} = \underline{0} \Rightarrow 2\underline{X}^T\underline{X}\,\underline{\beta}^* - 2\underline{X}^T\underline{y} = \underline{0}.$$

Therefore, $\underline{\beta}^* = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y}$.

observe here the pseudo inverse of the long matrix $\underline{X}$. In other words,

$$\underline{X}^+ = (\underline{X}^T\underline{X})^{-1}\underline{X}^T.$$

$\underline{\beta} = [\beta_1\ \beta_2\ \cdots\ \beta_k]^T$ characterizes the required (regression) line.

Scanned with CamScanner

EE306: Signals and Systems II

Lecture 11

Linear programming

* It refers to constrained optimization of a linear function.

→ Problem formulation:

minimize $\underline{c}^T\underline{x}$

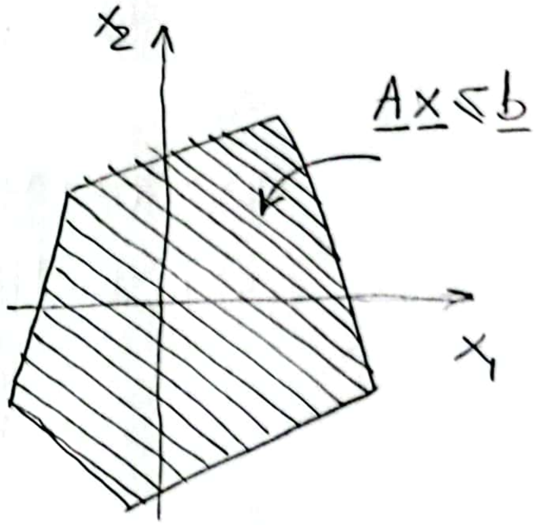
subject to $\underline{A}\underline{x} \leq \underline{b}$.

Q What happens if unconstrained?

EX1 Consider the following optimization problem: minimize $\underline{c}^T\underline{x}$

subject to $\underline{A}\underline{x} \leq \underline{b}$,

where $\underline{c}$ and $\underline{x}$ are in $\mathbb{R}^{2\times1}$, while $\underline{A}$ is in $\mathbb{R}^{5\times2}$, and b is in $\mathbb{R}^{5\times1}$. The constraint region is shaded in the figure.



Suggest where $\underline{x}^*$ that minimizes $\underline{c}^T\underline{x}$ should be.

Sol. Consider the line $c_1x_1 + c_2x_2 = a$.

As a gets smaller, this line keeps moving in a specific direction until it hits a corner, then leaves $\underline{A}\underline{x} \leq \underline{b}$ region $\Rightarrow$ $\underline{x}^*$ is at one of the 5 corners.

Q Can $\underline{x}^*$ be any point on a side?

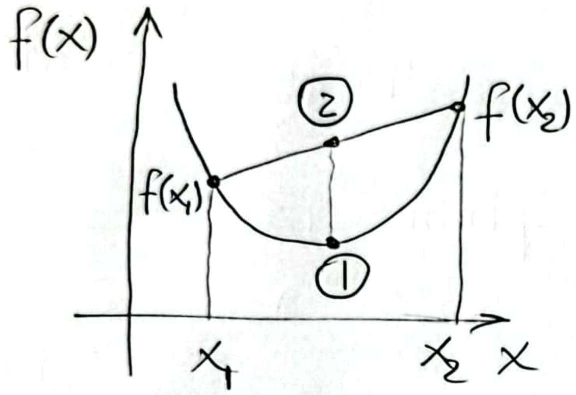
Convex-concave functions

* A function $f(x)$ is convex over $[a,b]$ if for every $(x_1, x_2) \in [a,b]$ and λ in $[0,1]$:

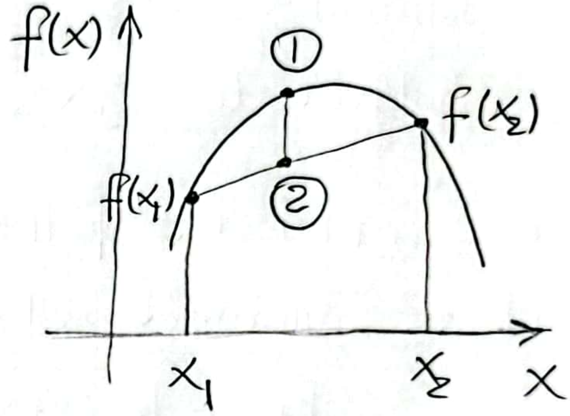
$$\textcircled{1}\ f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)\ \textcircled{2}.$$

Concave: $f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2)$.

→ strict convexity (concavity) if ineq. is strict.



convex function



concave function

* If $f_i(x)$, $i = 1, 2, \ldots, n$ are convex, then

$$g(x) = \sum_{i=1}^{n} c_i f_i(x), \quad c_i \geq 0, \ \forall i \text{ is also convex.}$$

<u>Proof</u> $g(\lambda x_1 + (1-\lambda)x_2) = \sum_{i=1}^{n} c_i f_i(\lambda x_1 + (1-\lambda)x_2)$

$$\leq \lambda \sum_{i=1}^{n} c_i f_i(x_1) + (1-\lambda) \sum_{i=1}^{n} c_i f_i(x_2)$$

$$= \lambda g(x_1) + (1-\lambda) g(x_2) \Rightarrow \text{convex.} \ \blacksquare$$

* If f has a second derivative that is non-negative (non-positive) over an interval, then the function is convex (concave) over that interval.

Q Is the linear fn convex or concave?

→ If f is convex, then $-f$ is concave.

Ex2 Test the convexity / concavity of :

(a) $f(x) = x^2$. (b) $f(x) = \log_e x = \ln x$.

Sol. (a) $f'(x) = 2x$, $f''(x) = 2 > 0 \Rightarrow$ convex.

(b) $f'(x) = \frac{1}{x}$, $f''(x) = \frac{-1}{x^2} < 0$ for its domain $x > 0$

$\Rightarrow$ concave.

* Everything also holds if $\underline{x}$ is a point in an n-dim. space (vector w.r.t. $\underline{0}$) instead of a point on a line.

Duality in optimization

* Standard form problem :

minimize $f_0(\underline{x})$

subject to $f_i(\underline{x}) \leq 0$, $i = 1, 2, -, m$

$h_i(\underline{x}) = 0$, $i = 1, 2, -, p$.

f_0 : objective fn

f_i, h_i, $i > 0$: const. fn's

variable $\underline{x} \in \mathbb{R}^n$, its domain is D, $f_0(\underline{x}^*) = p^*$.

→ Lagrangian $L(\underline{x}, \underline{\lambda}, \underline{\nu}) = f_0(\underline{x}) + \sum_{i=1}^{m} \lambda_i f_i(\underline{x}) + \sum_{i=1}^{p} \nu_i h_i(\underline{x})$

is a weighted sum of objective and const. fn's.

→ Lagrange dual fn : $g(\underline{\lambda}, \underline{\nu}) = \underbrace{\inf_{\underline{x}}}_{\text{(greatest lower bound)}} L(\underline{x}, \underline{\lambda}, \underline{\nu})$.

g is a concave fn (why?).

$\rightarrow$ If $\lambda_i \geqslant 0$, for all i, $g(\underline{\lambda}, \underline{\nu}) \leqslant f_0(\underline{x}) \leqslant p^*$.

Thus, $g(\underline{\lambda}, \underline{\nu})$ serves as a lower bound on p^*.

<u>Ex3</u> Minimize $\underline{x}^T \underline{x}$ subject to $\underline{A}\,\underline{x} = \underline{b}$.

<u>Sol.</u> $L(\underline{x}, \underline{\nu}) = \underline{x}^T \underline{x} + \underline{\nu}^T (\underline{A}\underline{x} - \underline{b})$.

$\Rightarrow \nabla_{\underline{x}} L(\underline{x}, \underline{\nu}) = 2\underline{x} + \underline{A}^T \underline{\nu}$

$\Rightarrow$ At $\underline{x} = \frac{-1}{2} \underline{A}^T \underline{\nu}$, $L(\underline{x}, \underline{\nu})$ is minimized

$\Rightarrow g(\underline{\nu}) = L(\frac{-1}{2} \underline{A}^T \underline{\nu}, \underline{\nu}) = -\frac{1}{4} \underline{\nu}^T \underline{A}\,\underline{A}^T \underline{\nu} - \underline{\nu}^T \underline{b}$.

Note that g is a concave fn and $g(\underline{\nu}) \leqslant p^*$.

<u>Ex4</u> Minimize $\underline{c}^T \underline{x}$ subject to $\underline{A}\,\underline{x} = \underline{b}$, $x_i \geqslant 0$ $\forall i$.

<u>Sol.</u> $L(\underline{x}, \underline{\lambda}, \underline{\nu}) = \underline{c}^T \underline{x} - \underline{\lambda}^T \underline{x} + \underline{\nu}^T (\underline{A}\underline{x} - \underline{b})$

$= -\underline{b}^T \underline{\nu} + (\underline{c} + \underline{A}^T \underline{\nu} - \underline{\lambda})^T \underline{x}$

$$\Rightarrow g(\underline{\lambda}, \underline{\nu}) = \inf_{\underline{x}} L(\underline{x}, \underline{\lambda}, \underline{\nu}) = \begin{cases} -\underline{b}^T \underline{\nu}, & \underline{A}^T \underline{\nu} - \underline{\lambda} + \underline{c} = \underline{0} \\ -\infty, & \text{o.w.} \end{cases}$$

Note that g is linear (also concave)

and $g(\underline{\nu}) = -\underline{b}^T \underline{\nu} \leqslant p^*$ if $\underline{A}^T \underline{\nu} + \underline{c} = \underline{\lambda} \geqslant \underline{0}$.

[Notation $\underline{x} \geqslant \underline{y}$ if and only if $x_i \geqslant y_i$, $\forall i$,
and the same for $\underline{x} \leqslant \underline{y}$, $\underline{x} > \underline{y}$, $\underline{x} < \underline{y}$.]