CPSC 536F Dec 1, 2025

Last time:

Poker:

Alice must choose

$p_1, p_2, \ldots, p_{52}$ so that

$$\min(\ell_1(\vec{p}), \ell_2(\vec{p}))$$

is maximized, ℓ_1, ℓ_2 are affine linear, i.e.

maximize z s.t.

$$z \le \ell_1(\vec{p})$$
$$z \le \ell_2(\vec{p})$$

$\forall i \in [52]$ $p_i \le 1$

and

$$p_1, p_2, \ldots, p_{52}, z \ge 0$$

Why take Math 340 (see my 2015-16 notes)?

<u>Great news</u>:

53 variables, 54 inequalities,

$0 < p_i < 1 \Rightarrow$ 2 vars (one p_i, one slack) $\neq 0$

$0 = p_i$ or $p_i = 1 \Rightarrow$ 1 var $\neq 0$

Either $z = \ell_1(\vec{p})$ or $z = \ell_2(\vec{p})$

$\Rightarrow$ 1 slack var there is 0.

$\boxed{\text{All but one } p_i \text{ must be } 0 \text{ or } 1}$

<u>Great news</u>: $p_1 \le p_2 \le \dots \le p_{52}$

<u>Great news</u>: "Vertices" of $0 \le p_1 \le \dots \le p_{52} \le 1$

are

$$\left.\begin{array}{cccc} p_1 & \le \dots & \le & p_{52} \\ 1 & \dots & 1 & 1 \\ 0 & 1 \ \dots & 1 & 1 \\ \vdots & & & \\ 0 & 0 \ \dots & & 0 \end{array}\right\} \text{53 vertices}$$

Great news: Duality $\Rightarrow$ can look at Betty's point of view! One decision:

call: prob q_1

fold: prob $1-q_1$

Hence only 1 variable !!!

Today:

- Ridge Regression + Lagrange Multipliers + Duality

- Kernel Interpolation (needs positive definite kernel function ...)

Recall: $X, \vec{y}$ given data points

$$(*) \quad \min_{\vec{w} \in \mathbb{R}^d} \|\underbrace{X\vec{w}}_{\text{model to approximate } \vec{y}} - \vec{y}\|^2 + \lambda \|\vec{w}\|^2$$

Trick: Equivalent to

$$(**) \quad \min_{\substack{\vec{z} \in \mathbb{R}^n \\ \vec{w} \in \mathbb{R}^d}} \left(\|z\|^2 + \lambda \|w\|^2 \right)$$

s.t.

$$\vec{z} = X\vec{w} - \vec{y}$$

$X \quad \vec{y}$ given

Lagrange multipliers: $f: \mathbb{R}^d \to \mathbb{R}$

$g: \mathbb{R}^d \to \mathbb{R}^a$

$$\min f(\vec{x})$$

$$\text{s.t. } \vec{g}(\vec{x}) = \vec{0} \quad \left(\vec{h}(\vec{x}) \leq \vec{0} \right)$$

(Can add $h(\vec{x}) \leq 0$ inequalities, but we don't need to here...)

w/o $g(\vec{x}) = 0$:

$$\nabla f(\vec{x}) = 0$$

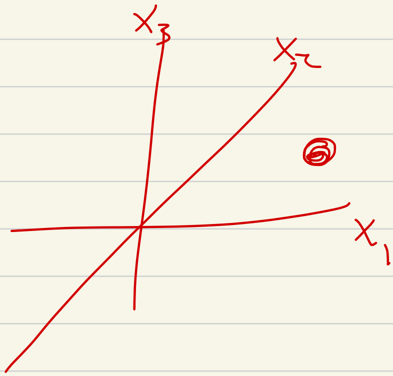
with $g(\vec{x}) = 0$

$$\nabla f(\vec{x}) + \alpha^T \nabla g(\vec{x}) = 0$$

$$\rightsquigarrow L(x, \vec{\alpha}) = f(\vec{x}) + \alpha^T g(\vec{x})$$

-5.1-

$f : \mathbb{R}^d \to \mathbb{R}$ $\quad$ (local) $\begin{pmatrix} \max \\ \min \end{pmatrix} f$

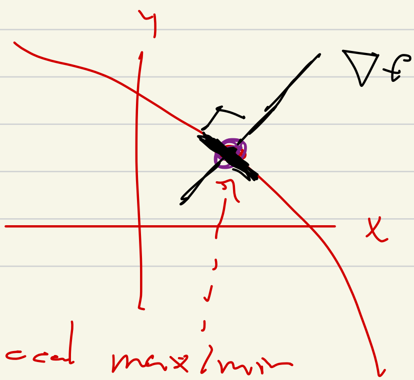


$\nabla f(\vec{x}) = 0$

$$f(\vec{w}) = \|X\vec{w} - \vec{y}\|^2 + \lambda \|\vec{w}\|^2$$

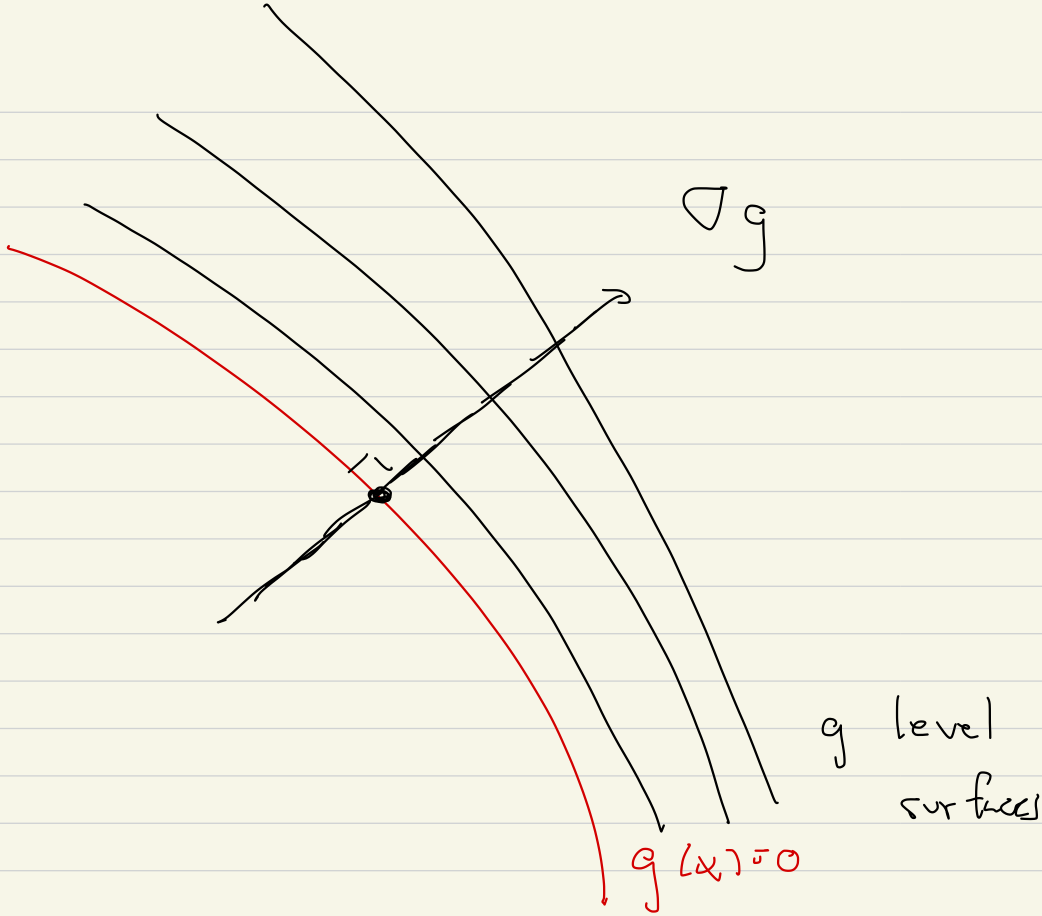
$\nabla f = 0 \Rightarrow$ normal eqs

$\mathbb{R}^3$



$\nabla f \perp$ direct where $g = 0$, i.e. tangent

$g(x,y) = 0$



$\nabla f \perp$ tangent of $g(x)=0$

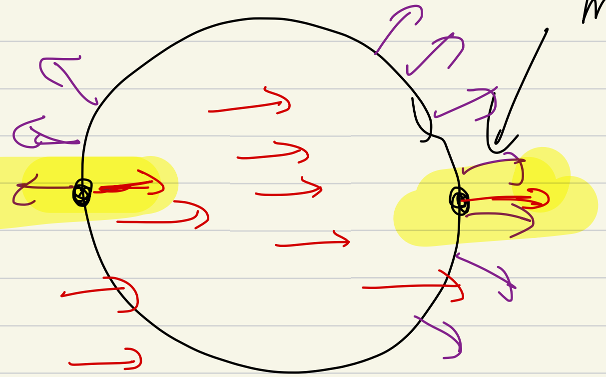
same

∇f proportional to ∇g

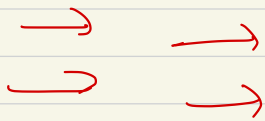
$x^2+y^2=1$

max $f(x,y)=x$

min $f(x,y)=x$



$\nabla f=(1,0)$



∇g, $g(x,y)=x^2+y^2-1$

$$\shortparallel$$

$$(2x,2y)$$

$$\nabla f + \vec{\alpha}\cdot\nabla \vec{g} = 0$$

In optimization:

$$\min \; f(\vec{x})$$

s.t.

$$\vec{g}(\vec{x}) = \vec{0}$$

this can be $\alpha = 1, 2, \ldots$ equations

Lagrangian:

$$L = L(\vec{x}, \vec{\alpha}) = f(\vec{x}) + \vec{\alpha} \cdot \vec{g}(\vec{x})$$

$$\nabla_{\text{w.r.t. } \vec{x}} L = \nabla f + \vec{\alpha} \cdot \vec{\nabla} g$$

at local min or max of

f s.t. $g = 0$.

$$\underset{\vec{x} \in \mathbb{R}^d}{\overset{\text{Betty}}{\min}} \underbrace{\left(\underset{\vec{\alpha} \in \mathbb{R}^a}{\overset{\text{Alice}}{\max}} L(x, \alpha) \right)}_{\substack{\text{fix } \vec{x} : \text{ if } \\ \vec{g}(\vec{x}) \neq \vec{0} \\ \text{then } \vec{\alpha} \cdot g(\vec{x}) \\ \text{can be made arbitrarily large}}}$$

$\overset{??}{=}$

Alice wants L as large as poss

Betty " " " small as possible

— 5.9.1 —

$$L(\vec{x}, \vec{\alpha}) = f(\vec{x}) + \vec{\alpha} \cdot \vec{g}(\vec{x})$$

For any $\vec{x}$:

$\vec{g}(\vec{x}) \not\leq \vec{0} \Rightarrow L(\vec{x}, \vec{\alpha})$ can be made $\rightarrow +\infty$ for appropriate $\vec{\alpha}$

$\vec{g}(\vec{x}) = \vec{0} \Rightarrow L(\vec{x}, \vec{\alpha}) = f(\vec{x})$

So

$$\min_{\vec{x} \in \mathbb{R}^d} \left(\max L \right)$$

$$= \min_{\substack{\vec{x} \in \mathbb{R}^d \\ \vec{g}(\vec{x}) = \vec{0}}} f(\vec{x})$$

Rem: Example of why L is not really canonical

$\vec{g}(\vec{x}) = 0 \iff 3\, g(\vec{x}) = 0$

iff

$(\vec{g}(\vec{x}))^5 = 0$

So

$$\min_{g(\vec{x})=0} \_\_ = \min_{(g(\vec{x}))^5=0}$$

$\Rightarrow$

$$L(\vec{x}, \vec{\alpha}) = f(\vec{x}) + \vec{\alpha}\,(\vec{g}(x))^5$$

$$L = f(\vec{x}) + \vec{\alpha}\; 3(\vec{g}(x))$$

$$\min \quad \max \quad L = \min_{\substack{\vec{x} \in \mathbb{R}^d \\ \vec{g}(\vec{x}) = \vec{0}}} f(\vec{x})$$

Now... in good situations:

$$\max_{\vec{\alpha} \in \mathbb{R}^a} \min_{\vec{x} \in \mathbb{R}^d} L = \min_{\vec{x} \in \mathbb{R}^d} \max_{\vec{\alpha} \in \mathbb{R}^a} L$$

if =, i.e.
duality gap
= 0

then these are
equivalent problems!!!

This doesn't always happen... we are
lucky when it does...

Consider

$$\min_{\vec{x} \in \mathbb{R}^d} \left(\max_{\vec{\alpha} \in \mathbb{R}^a} L(x, \vec{\alpha}) \right)$$

and

$$\max_{\vec{\alpha}} \left(\min_{\vec{x}} L \right)$$

Kernel interpolation + positive definiteness?

Data: $(\vec{x}_1, y_1), (\vec{x}_2, y_2), \ldots, (\vec{x}_n, y_n)$

Want

$f(x)$ s.t.

(1) $f(\vec{x}_i) = y_i$, and

(2) $f(\vec{x}) = \sum_{i=1}^{n} \alpha_i k(x, x_i)$

$\Longrightarrow$ So

$$K \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

where ...