Convex Optimization Problems: Duality

Consider

$$\text{Minimize} \quad f_0(\underline{x})$$

ST

$$f_i(\underline{x}) \leq 0 \quad ; \quad 1 \leq i \leq m$$

$$h_i(\underline{x}) = 0 \quad ; \quad 1 \leq i \leq k$$

Assumption: $\exists$ at least one feasible point

Define the Lagrangian $\quad L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$

$$L(\underline{x}, \underline{\lambda}, \underline{v}) = f_0(\underline{x}) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{k} v_i g_i(\underline{x})$$

$\underline{\lambda}, \underline{v}$ are called Lagrange multipliers

Lagrange dual function

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = -\underbrace{\sup_{x \in \mathcal{D}} \left(\underbrace{-L(x, \lambda, \nu)}_{\text{linear fn of } (\lambda, \nu) \text{ for each } x, \text{ hence convex}} \right)}_{\text{Convex}}$$

$$\mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom}(f_i) \bigcap_{j=1}^{k} \operatorname{dom}(h_j)$$

(Not constraint set!)

<u>Property 1</u>: $g(\lambda, \nu)$ is concave

Property 2 : Let $p^* = \inf f_0(x)$

st $f_i(x) \le 0 \quad 1 \le i \le m$

$h_j(x) = 0 \quad 1 \le j \le k$ } constraint set C

Then, $g(\lambda, \nu) \le p^* \quad \forall (\lambda, \nu)$ with $\lambda \succeq 0$

If $x \in C$, then $f_i(x) \le 0 \quad \forall i$

$h_i(x) = 0 \quad \forall i$

$\Rightarrow L(x, \lambda, \nu) \le f_0(x)$ for $\lambda \succeq 0$.

$\Rightarrow g(\lambda, \nu) = \inf\limits_{x \in \mathcal{D}} L(x, \lambda, \nu) \le \inf\limits_{x \in C} L(x, \lambda, \nu)$

$\le \inf\limits_{x \in C} f_0(x) = p^*$.

Eg

$f_0(x) = x^T x$

ST $\quad Ax = b$

Assume $Ax = b$ has at least one solution

$$L(x, v) = \underbrace{x^T x + v^T (Ax - b)}_{\text{convex}}$$

$$\frac{\partial L}{\partial x} = 0 \Rightarrow \quad 2x + A^T v = 0$$

$$x = -A^T v$$

$$g(v) = \inf_{x \in \mathcal{D}} L(x, v) = -\frac{1}{4} v^T A A^T v - v^T b$$

What is $\max_{v} g(v)$? $\quad$ Compare this with $\inf_{Ax=b} f_0(x)$.

Take an example: $\quad A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

Problem: Minimize $x_1^2 + x_2^2 = f_0(x)$

ST $x_1 + x_2 = 1$

$x_1 + x_2 - 1 = 0$

$$g(\nu) = \inf_{x' \in \mathbb{R}^2} \left[f_0(x) + \nu(x_1 + x_2 - 1) \right]$$

$$= -\frac{\nu^2}{2} - \nu$$

Minimize $x_1^2 + (1 - x_1)^2$ $\quad p^* = \frac{1}{2}$

Maximize $-\frac{\nu^2}{2} - \nu$ $\quad d^* = \frac{1}{2}$

Eg : Min $f(x) = x^3$

ST $x \geq 1 \Rightarrow -x+1 \leq 0$

$$L(x, \lambda) = x^3 + \lambda(-x+1)$$

$$g(\lambda) = \inf_x \left\{ x^3 + \lambda(-x+1) \right\}$$

$$= -\infty$$

Redefine

$$f(x) = \begin{cases} x^3 & \text{for } x \geq 0 \\ +\infty & \text{for } x < 0 \end{cases}$$

$$L(x, \lambda) = f(x) + \lambda(-x+1) = \begin{cases} x^3 + \lambda(-x+1) & \text{if } x \geq 0 \\ \infty & \text{if } x < 0 \end{cases}$$

$$g(\lambda) = \inf_{x \in \mathbb{R}} L(x, \lambda)$$

$$= \begin{cases} \lambda\left(1 - \frac{2}{3}\sqrt{\frac{\lambda}{3}}\right) & \lambda \geq 0 \\ \end{cases}$$

$$\sup_{\lambda \geq 0} g(\lambda) =$$

Original problem (Primal)

$$\begin{aligned} \text{Minimize} \quad & f_0(\underline{x}) \\ \text{ST} \quad & f_i(\underline{x}) \leq 0 \qquad 1 \leq i \leq m \\ & h_j(\underline{x}) = 0 \qquad 1 \leq j \leq k \end{aligned}$$

Optimum value p^*

Dual problem

$$\begin{aligned} \text{Maximize} \quad & g(\underline{\lambda}, \underline{\nu}) \\ \text{ST} \quad & \underline{\lambda} \succcurlyeq 0 \end{aligned}$$

Optimum value d^*

Weak duality : $d^* \leq p^*$

Weak duality : $d^* \le p^*$

Strong duality : $d^* = p^*$

Suppose strong duality holds.

$$p^* = d^* = \sup_{\substack{\lambda \succeq 0 \\ \nu}} g(\lambda, \nu)$$

$$= \sup_{\substack{\lambda \succeq 0 \\ \nu}} \inf_{x} L(x, \lambda, \nu)$$

Even if strong duality does not hold,
can get a lower bound.

$$f(x) = x^2 \quad \text{st} \quad x \geq 1$$

$$L(x, \lambda) = x^2 + \lambda(-x+1)$$

Eg 1

$$f(x,y) = \begin{cases} e^{-x} & \text{for } x \in \mathbb{R},\ y > 0 \\ \infty & \text{else} \end{cases}$$

$$\text{st} \quad \frac{x^2}{y} \leq 0 \qquad\qquad p^* = 1$$

$$L(x, y, \lambda) = f(x,y) + \lambda \frac{x^2}{y}$$

$$= \begin{cases} e^{-x} + \lambda x^2/y & \forall\ x \in \mathbb{R},\ y > 0 \\ \infty & \forall\ y \leq 0 \end{cases}$$

$$g(\lambda) = \inf_{x,y} L(x, y, \lambda)$$

$$= \begin{cases} 0 & \lambda \geq 0 \\ -\infty & \lambda < 0 \end{cases}$$

$y = x^3$
$x \to \infty$
$\Rightarrow L(x, y, \lambda) \to 0$

$y > 0$
$x \to \infty$
$\Rightarrow L(x, y, \lambda) \to -\infty$

$d^* = 0$

Weak duality holds, Strong duality does not hold

Strong duality holds if

f is convex

$\exists\, x \in \text{relint}(\mathcal{D})$ st

$f_i(x) < 0 \quad \forall i$

$h_i(x) = 0 \quad \forall i$

$f_i(x) \leq 0$

$\bar{f}(x) < 0$

$\bar{h}(x) = 0$

$$g(\underline{\lambda}, \underline{\nu}) = \inf_{\underline{x} \in \mathcal{D}} \left(f_0(x) + \sum_i \lambda_i f_i(x) + \sum_j \nu_j h_j(x) \right)$$

Minimize $f_0(\underline{x})$

$$A\underline{x} - \underline{b} \preceq 0$$

$$C\underline{x} - \underline{d} = 0$$

$$g(\underline{\lambda}, \underline{\nu}) = \inf_{\underline{x} \in \mathcal{D}} \left(f_0(\underline{x}) + \underline{\lambda}^T (A\underline{x} - \underline{b}) + \underline{\nu}^T (C\underline{x} - \underline{d}) \right)$$

$$= -\lambda^T b - \nu^T d + \inf \left(f_0(x) + \lambda^T (A x) + \nu^T (C x) \right)$$

$$= -\lambda^T b - \nu^T d + \inf_x \left(f_0(x) + x^T \left(A^T \lambda + C^T \nu \right) \right)$$

$$g(\lambda, \nu) = -\lambda^T b - \nu^T d - \sup \left(-x^T \left(A^T \lambda + C^T \nu \right) - f_0(x) \right)$$

$$= -\lambda^T b - \nu^T d - \sup_{x} \left(x^T \left(-A^T \lambda - C^T \nu \right) - f_0(x) \right)$$

$$g(\lambda, \nu) = -\lambda^T b - \nu^T d - f^* \left(-A^T \lambda - C^T \nu \right)$$

$\downarrow$

conjugate

Original problem (Primal)

$$\text{Minimize} \quad f_0(\underline{x})$$

$$\text{ST} \quad f_i(\underline{x}) \leq 0 \qquad 1 \leq i \leq m$$

$$h_j(\underline{x}) = 0 \qquad 1 \leq j \leq k$$

Optimum value p^*

Dual problem

$$\text{Maximize} \quad g(\underline{\lambda}, \underline{\nu})$$

$$\text{ST} \quad \underline{\lambda} \succeq 0$$

Optimum value d^*

$$\boxed{\begin{array}{ll} \text{Weak duality:} & d^* \leq p^* \\ \text{Strong duality:} & d^* = p^* \end{array}}$$

If strong duality holds,

$$p^* = d^* = \sup_{\substack{\underline{\lambda} \succeq 0 \\ \underline{\nu} \in \mathbb{R}^k}} \inf_{\underline{x} \in \mathcal{D}} L(\underline{x}, \underline{\lambda}, \underline{\nu})$$

Consider a minimization problem

$$\min_{\underline{x} \in K} f(\underline{x})$$

convex

convex

$$= \text{Minimize} \quad f(\underline{x}) + \mathbb{I}_K(\underline{x})$$

$$\mathbb{I}_K(x) = \begin{cases} 0 & \text{if } \underline{x} \in K \\ \infty & \text{else} \end{cases}$$

$x_1 \; x_2 - \alpha \qquad x_1 \notin K, \quad x_2 \notin K$

$$f(\alpha x_1 + (1-\alpha) x_2) \leq$$

$$\alpha f(x_1) + (1-\alpha) f(x_2)$$

$$F(\underline{x}_1, \underline{x}_2) = f(\underline{x}_1) + \mathbb{1}_K(\underline{x}_2)$$

Suppose

$$F(\underline{x}) = f_1(\underline{x}) + f_2(A\underline{x}) \longrightarrow$$ original function to minimize

where f_1 & f_2 are convex

$$F(\underline{x}_1, \underline{x}_2) = f_1(\underline{x}_1) + f_2(\underline{x}_2) \longrightarrow$$ constrained minimization

$$\text{ST} \quad \underline{x}_2 = A\underline{x}_1$$

$$L(\underline{x}_1, \underline{x}_2, \underline{v}) = f_1(\underline{x}_1) + f_2(\underline{x}_2) + \underline{v}^T(\underline{x}_2 - A\underline{x}_1)$$

$$g(\underline{v}) = \inf_{\underline{x}_1, \underline{x}_2} \left(f_1(\underline{x}_1) + f_2(\underline{x}_2) + \underline{v}^T \underline{x}_2 - v^T A x_1 \right)$$

$$= \inf_{\underline{x}_1} \left(f_1(\underline{x}_1) - \underline{v}^T A x_1 \right) + \inf_{\underline{x}_2} \left(f_2(x_2) + v^T x_2 \right)$$

$$= - \sup_{\underline{x}_1} \left((A^T v)^T \underline{x}_1 - f_1(x_1) \right)$$

$$- \sup \left(-v^T x_2 - f_2(x_2) \right)$$

$$= - f_1^*(A^T \underline{v}) - f_2^*(-v)$$

Fenchel duality: When $\mathfrak{b}$

$$\inf_{\underline{x}} \left(f_1(\underline{x}) + f_2(A\underline{x}) \right) = \sup_{\underline{v}} \left(- f_1^*(A^T\underline{v}) - f_2^*(-\underline{v}) \right)$$

Convex optimization problem

convex

Minimize $f_0(x)$

ST $f_i(x) \leq 0$

$h_i(x) = 0$

Problem Quadratic programming with quadratic constraint

$$\text{Minimize} \quad \underline{x}^T A \underline{x} + 2\underline{b}^T \underline{x}$$

$$\text{ST} \quad \|\underline{x}\|^2 \leq 1$$

$$A \in S^n$$

$$L(\underline{x}, \lambda) = \underline{x}^T A \underline{x} + 2\underline{b}^T \underline{x} + \lambda(\|\underline{x}\|^2 - 1)$$

$$= x^T A x + 2b^T x + \underbrace{\lambda(x^T x}_{x^T(\lambda I)x} - 1)$$

$$= x^T(A + \lambda I)x + 2b^T x - \lambda$$

$$g(\lambda) = \inf_{x} \left[x^T (A + \lambda I) x + 2b^T x - \lambda \right]$$

$$= \begin{cases} -b^T (A+\lambda I)^{-1} b - \lambda & A + \lambda I \succcurlyeq 0 \\ -\infty & A + \lambda I \not\succcurlyeq 0 \end{cases}$$

$\Downarrow$

If $A + \lambda I$ is not PSD,

$\exists\, x_0$ st $x_0^T (A + \lambda I) x_0 < 0$

& $2b^T x_0 < 0$

Take $x = \alpha x_0$ & $\alpha \to \infty$

$$\nabla_{\underline{x}} L(\underline{x}, \lambda) = \nabla_{\underline{x}}\left(\underline{x}^T (A+\lambda I)\underline{x} + 2b^T\underline{x} - \lambda \right)$$

$$= 2(A+\lambda I)\underline{x} + 2b = 0$$

$$\Rightarrow \underline{x} = -(A+\lambda I)^{\dagger} b$$

Dual optimization problem

$$\begin{aligned} &\text{Maximize } g(\lambda) \\ &\text{st } \lambda \geq 0 \end{aligned}$$

$$\equiv \quad \begin{aligned} &\text{Maximize } -b^T(A+\lambda I)^{-1} b^T - \lambda \\ &\text{st } \lambda \geq 0 \\ &\quad A+\lambda I \succeq 0 \end{aligned}$$

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} 1/2 \\ 1/3 \end{bmatrix} \qquad A + \lambda I = \begin{bmatrix} 2+\lambda & 0 \\ 0 & \lambda - 1 \end{bmatrix}$$

$$f(\underline{x}) = 2x_1^2 - x_2^2 + x_1 + \frac{2}{3}x_2$$

$$\text{ST} \quad x_1^2 + x_2^2 \leq 1$$

$$g(\lambda) = -b^T (A + \lambda I)^{-1} \underline{b} - \lambda \qquad \text{ST} \quad A + \lambda I \succeq 0$$

$$= -b^T \begin{bmatrix} \frac{1}{2+\lambda} & 0 \\ 0 & \frac{1}{\lambda - 1} \end{bmatrix} \underline{b} - \lambda$$

$$= -\frac{1}{4}\,\frac{1}{2+\lambda} - \frac{1}{9}\,\frac{1}{\lambda - 1} - \lambda \qquad \text{ST} \quad \lambda \geq 1$$

Standard form LP

Minimize $\quad c^T x$

ST $\quad Ax = b$

$\quad x \geq 0$

$$L(x, \lambda, \nu) = c^T x - \lambda^T (x) + \nu^T (Ax - b)$$

$$g(\lambda, \nu) = \inf_x \left[c^T x - \lambda^T (x) + \nu^T (Ax - b) \right]$$

$$= \inf_x \left[x^T (A^T \nu - \lambda + c) - \nu^T b \right]$$

$$= \begin{cases} -\nu^T b & \text{if } A^T \nu - \lambda + c = 0 \\ -\infty & \text{else} \end{cases}$$

Dual program: Maximize $-\underline{v}^T \underline{b}$

ST $\underline{\lambda} \geq 0$

$A^T v - \lambda + c = 0$

$\equiv$ Maximize $-\underline{v}^T b$

ST $A^T v + c \geq 0$

Strong duality holds for all LPs.

Entropy Maximization

For any pmf p, over $\{1, 2, \ldots n\}$

$$H(p) = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i}$$

Goal: Minimize $-H(p)$

st $\underbrace{Ap \leq b}_{\text{Moment constraints}}$

Minimize $-H(p)$ $\qquad p = [p_1 \; p_2 \; p_3]$

st $\quad p_1 + 2p_2 + 3p_3 \leq 1.5$

$\quad p_1 + p_2 + p_3 = 1, \quad p \geq 0$

$x \log x$ is convex $\Rightarrow$ This is a convex optimization problem

$$L(p, \lambda, \underline{v}) = \sum_{i=1}^{n} p_i \log p_i + \underline{\lambda}_1^T(-p) + \underline{\lambda}_2^T(Ap - b)$$

$$+ v(p_1 + p_2 + p_3 - 1)$$

$$g(\lambda, \underline{v}) = \Bigg\{$$

$$\frac{\partial L}{\partial p_i} =$$

Optimality conditions

minimize $f_0(x)$

st $f_i(x) \leq 0 \quad i = 1, 2, \ldots m$

$h_j(x) = 0 \quad j = 1, \ldots k$

Assumption: problem is feasible

$\exists\ x^*$ for which $f_0(x^*) = p^*$

& feasible

$$d^* = \sup_{\substack{\lambda \succeq 0 \\ \nu \in \mathbb{R}^k}} g(\lambda, \nu)$$

Assume λ^* & ν^* achieve d^*.

$$p^* \geq g(\lambda, \nu) \quad \text{for any pr } \lambda \succeq 0,\ \nu \in \mathbb{R}^k$$

For any pr x satisfying constraints, (dual feasible)

$$f_0(x) - p^* \leq f_0(x) - g(\lambda, \nu)$$

$\Rightarrow$ x is an ϵ - approximate solution

$$\epsilon = \underbrace{f_0(x) - g(\lambda, \nu)}_{\text{Duality gap}}$$

Suppose we design an iterative algorithm that

produces $\underbrace{x^{(t)}}_{\text{primal feasible}}, \underbrace{\lambda^{(t)}, \nu^{(t)}}_{\text{dual feasible}}$

If $f_0(x^{(t)}) - g(\lambda^{(t)}, \nu^{(t)}) \leq \epsilon$

then $x^{(t)}$ is ϵ-close to the optimum

$$f_0(x) - p^* \leq \epsilon$$

$$f_0(x^*) = g(\lambda^*, v^*)$$ (Strong duality)

$$= \inf_{x} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{j=1}^{k} v_i^* h_i(x) \right\}$$

$$\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{j=1}^{k} v_i^* h_i(x^*)$$

$$x_0^* = \min f_0(x)$$
$$\text{s.t. } f_i(x) \leq 0$$
$$h_i(x) = 0$$

Need not be the same

$$\Rightarrow \quad 0 \leq \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{j=1}^{k} \nu_j^* h_j(x^*)$$

$\underline{\lambda}^* \geq 0$, $\underline{\nu}^* \in \mathbb{R}$, x^* is primal feasible

$$\sum_{j=1}^{k} \nu_j^* h_j(x^*) = 0 \qquad \text{since } h_j(x^*) = 0 \ \forall j$$

$$\Rightarrow \quad \lambda_i^* f_i(x^*) = 0 \quad \forall i$$

$$\Rightarrow \quad \lambda_i^* = 0 \ \text{ OR } \ f_i(x^*) = 0$$

Complementary slackness conditions

Suppose x^* solves the primal problem & λ^*, ν^* solves the dual problem

(and all f_i, h_i are differentiable)

1. $f_i(x^*) \leq 0 \quad \forall i$
2. $h_j(x^*) = 0 \quad \forall j$
3. $\lambda_i^* \geq 0 \quad \forall i$
4. $\lambda_i^* f_i^*(x) = 0 \quad \forall i$
5. $\nabla_x L(x^*, \lambda^*, \nu^*) = \nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) + \sum_j \nu_j^* \nabla h_j(x^*) = 0$

KKT Conditions

(Karush - Kuhn - Tucker)

$$L(x^*, \lambda^*, \nu^*) = f_0(x^*)$$

for any feasible x,

$$L(x, \lambda^*, \nu^*) = f_0(x) + \underbrace{\sum_{i=1}^{m} \lambda^*_i f_i(x)}_{\leq 0}$$

$$f_0(x^*) = L(x^*, \lambda^*, \nu^*) \leq L(x, \lambda^*, \nu^*)$$

$$\begin{array}{c} \shortparallel \\ g(\lambda^*, \nu^*) = \inf_x L(x, \lambda^*, \nu^*) \end{array}$$

If $f_0, f_1 - f_m$ are convex & $h_1 - h_k$ are affine, then KKT conditions are sufficient

If x^*, λ^*, ν^* satisfy KKT conditions,

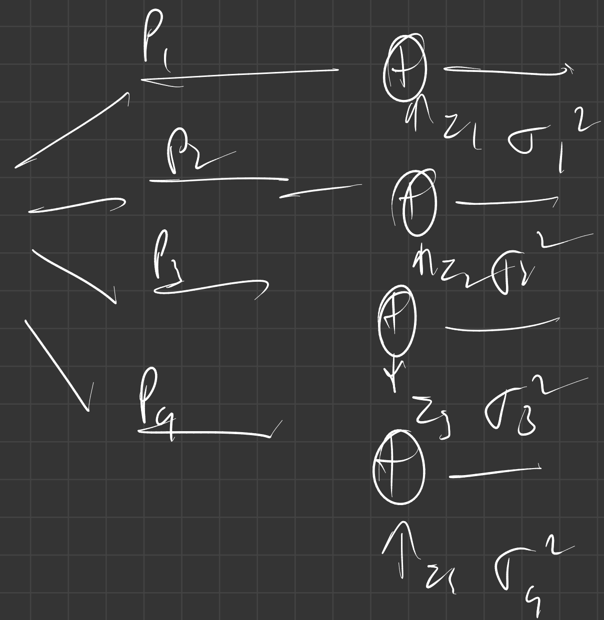
then

$\nabla L(x^*, \lambda^*, \nu^*) = 0 \Rightarrow x^*$ minimizes $L(x, \lambda^*, \nu^*)$

$$L(x^*, \lambda^*, \nu^*) = g(\lambda^*, \nu^*)$$
$$= f_0(x^*)$$

Zero duality gap $\Rightarrow$ x^* is the optimizer

Power allocation across channels



$$R_i = \frac{1}{2} \log\left(1 + \frac{P_i}{\sigma_i^2}\right)$$

$$R = \sum_{i=1}^{m} R_i$$

Power constraint:

$$\sum_{i=1}^{m} P_i \leq P$$

$$f(\underline{p}) = -\sum_{i=1}^{m} \log(\sigma_i^2 + p_i)$$

Minimize

Maximize

$$\sum_{i=1}^{m} p_i \leq P$$

$$\underline{p} \succeq 0$$

$$L(\underline{p}, \lambda) = -\sum_{i=1}^{m} \log(\sigma_i^2 + p_i) + \lambda_1\left(\sum_{i=1}^{m} p_i - P\right)$$

$$- \underline{\lambda}_2^T \underline{p}$$

$$① \quad -\frac{1}{\sigma_i^2 + p_i} + \lambda_1 - \lambda_{2i} = 0$$

$$\Rightarrow \quad \lambda_{2i} = \lambda_1 - \frac{1}{\sigma_i^2 + p_i}$$

② $\lambda_2 \geq 0 \Rightarrow \lambda_{2i} \geq 0 \quad \forall i$

$$\lambda_1 - \frac{1}{\sigma_i^2 + \rho_i} \geq 0$$

③ $\lambda_1 \geq 0$

$$\lambda_1 - \frac{1}{\sigma_i^2}$$

④ $\lambda_1 \left(\sum_{i=1}^{n} \rho_i - P \right) = 0$

$\lambda_2^T \rho = 0 \Rightarrow \lambda_{2i} \rho_i = 0$

$$\rho_i \left(\lambda_1 - \frac{1}{\sigma_i^2 + \rho_i} \right) = 0$$

$$\Rightarrow \quad P_i = 0 \quad \text{or} \quad \lambda_1 = \frac{1}{\sigma_i^2 + P_i}$$

$$P_i = \frac{1}{\lambda_1} - \sigma_i^2$$

$$P_i \geq 0 \quad \Rightarrow \quad P_i = \begin{cases} 0 & \text{if } \frac{1}{\lambda_1} - \sigma_i^2 < 0 \\ \frac{1}{\lambda_1} - \sigma_i^2 & \text{if } \frac{1}{\lambda_1} - \sigma_i^2 \geq 0 \end{cases}$$

$$P_i = \max\left\{0, \frac{1}{\lambda_1} - \sigma_i^2\right\}$$

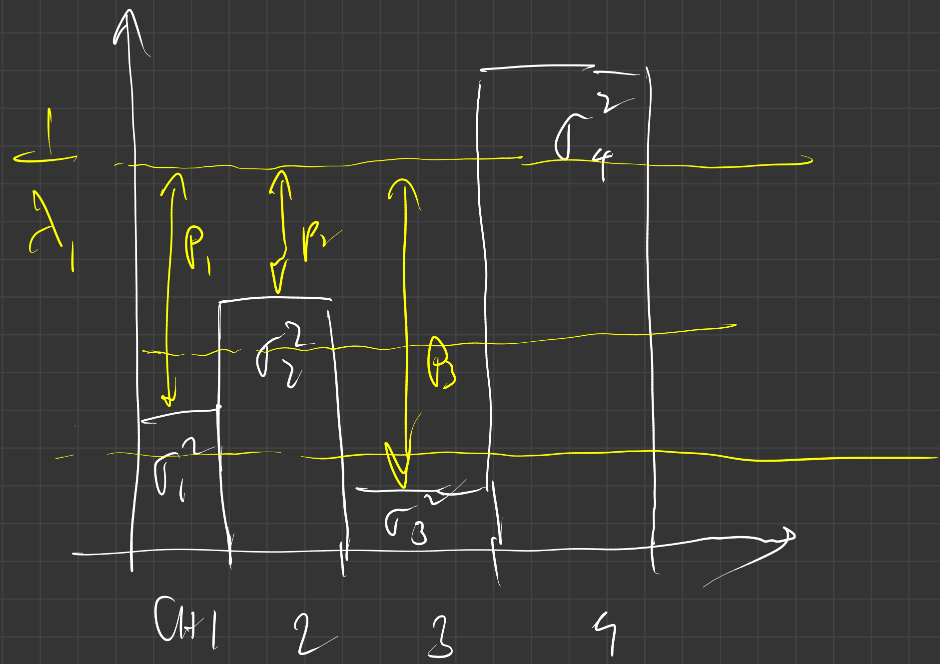
$$\frac{1}{\lambda_1} - \sigma_i^2 > 0 \qquad \lambda_{2i} = \lambda_1 - \frac{1}{P_i + \sigma_i^2} \geq 0$$

$$\frac{1}{\lambda_1} \geq \sigma_i^2$$

$$\frac{1}{\sigma_i^2} \geq \lambda_1 \qquad \lambda_1 - \frac{1}{\sigma_i^2} < 0$$

$$\sum_{i=1}^{n} p_i = p$$

$$\sum_{i=1}^{n} \max\left\{0, \frac{1}{\lambda_1} - \sigma_i^2\right\} = p$$



Waterfilling solution

CLASSIFICATION

Ground truth : S_1 S_2 disjoint

$S_1, S_2 \subseteq \mathbb{R}^n$

$S_1 \rightarrow$ label 0, $S_2 \rightarrow$ label 1

$(\underline{x}_1, y_1), (x_2, y_2)$ — — (x_m, y_m)

$$y_i = \begin{cases} 0 & \text{if } \underline{x}_i \in S_1 \\ 1 & \text{if } x_i \in S_2 \end{cases}$$

Goal: design f that predicts if $x \in S_1$ or $x \in S_2$

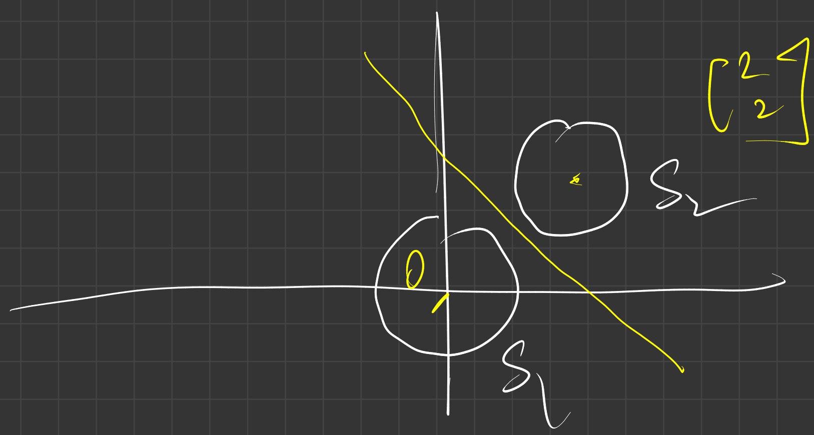
$$f(x) = \begin{cases} 0 & \text{if} \quad \underline{a}^T \underline{x} \leq b \\ 1 & \text{if} \quad \underline{a}^T \underline{x} > b \end{cases}$$

linear classifier

For eg, $S_1 = B(0, 1)$ $\quad S_2 = B(2\underline{1}, 1)$

$\downarrow$

all ones vector



Assume that S_1 & S_2 are separable.

$$\underbrace{x_1^{(0)} \;\cdots\; x_{k_1}^{(0)}}_{S_1} \qquad \underbrace{x_1^{(1)} \;\cdots\; x_{k_2}^{(1)}}_{S_2}$$

Minimize $\displaystyle\sum_{x_i^{(0)}:\ a^T x_i^{(0)} > b} 1 \;+\; \sum_{x_i^{(1)}:\ a^T x_i^{(1)} \le b} 1$

Finding a classifier is the same as solving

$$\min \quad 1$$

$$\text{ST} \quad \underline{a}^T \underline{x}_i^{(0)} \leq b \qquad i = 1, 2, \ldots, k_1$$

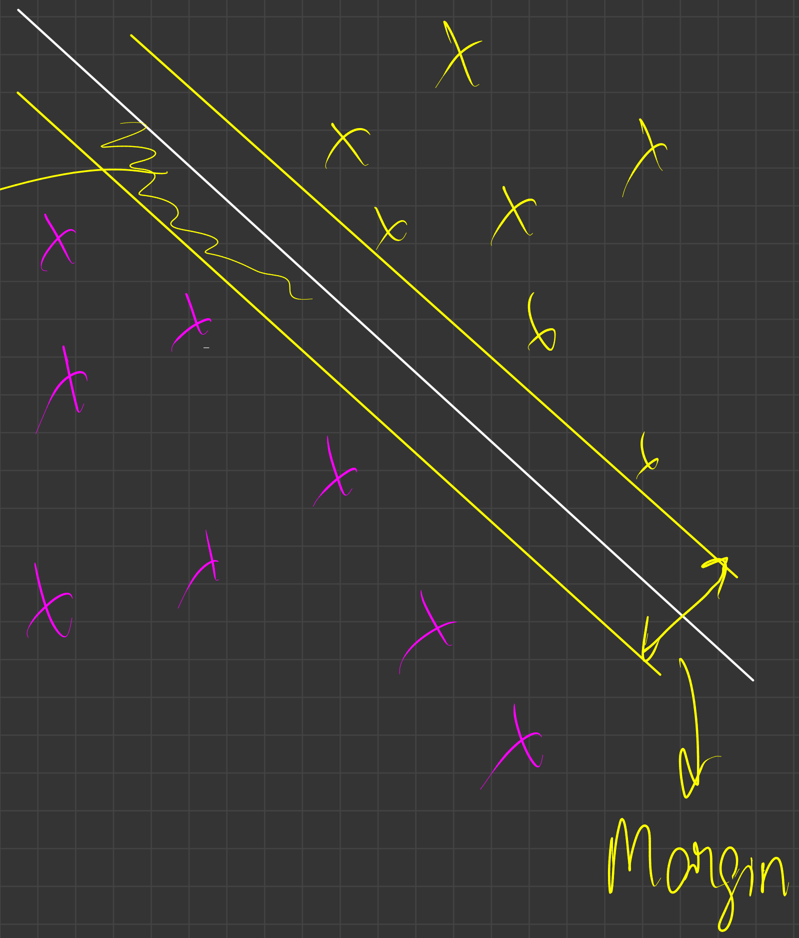
$$\underline{a}^T x_i^{(1)} \geq b \qquad i = 1, 2, \ldots, k_2$$

This is a feasibility problem

Goal:

Construct a maximum margin linear classifier

Buffer zone (contains no points $\underline{x}_i^{(0)}$ or $x_i^{(1)}$)



$$\min \ -t$$

$$\text{st} \quad \underline{a}^T \underline{x}_i^{(0)} \leq b - t \qquad i = 1, 2, \ldots, k_1$$

$$\underline{a}^T x_i^{(1)} \geq b + t \qquad i = 1, 2, \ldots, k_2$$

$$\|a\|^2 \leq 1$$

<u>Lagrangian:</u>

$$L(t, \underline{a}, b, \underline{\lambda}_1, \underline{\lambda}_2, \lambda_3) = -t + \sum_{i=1}^{k_1} \lambda_{1i} \left(\underline{a}^T \underline{x}_i^{(0)} - b + t\right)$$

$$+ \sum_{j=1}^{k_2} \lambda_{2j} \left(b + t - \underline{a}^T x_j^{(1)}\right)$$

$$+ \lambda_3 \left(\|\underline{a}\|^2 - 1\right)$$

$$= t\left[-1 + \sum_{i=1}^{k_1} \lambda_{1i} + \sum_{j=1}^{k_2} \lambda_{2j}\right]$$

$$+ b\left[-\sum_{i=1}^{k_1} \lambda_{1i} + \sum_{j=1}^{k_2} \lambda_{2j}\right]$$

$$+ \sum_{i=1}^{k_1} \lambda_{1i}\, a^T x_i^{(0)} + \sum_{j=1}^{k_2} (-\lambda_{2j})\, a^T x_j^{(1)}$$

$$+ \lambda_3 (\|a\|^2 - 1)$$

$$\inf_{t,b} L(\,) = \begin{cases} (\ ) & \sum_i \lambda_{1i} = \sum_j \lambda_{2j} = \frac{1}{2} \\ -\infty & \text{else} \end{cases}$$

$$\inf_{a} \sum_{i=1}^{k_1} \lambda_{1i}\, a^T x_i^{(0)} + \sum_{j=1}^{k_2} (-\lambda_{2j})\, a^T x_j^{(1)}$$

$$+ \lambda_3 (\|a\|^2 - 1)$$

$$g(\lambda_1, \lambda_2, \lambda_3) = \begin{cases} -\lambda_3 & \text{if } \left\| \sum_i \lambda_{1i} x_i^{(0)} - \sum_j \lambda_{2j} x_i^{(1)} \right\| \le \lambda_3 \\ -\infty & \text{else} \end{cases}$$

Dual optimization problem

pt in the convex hull of pts label 0

Maximize $-\lambda_3$

ST

$$\frac{1}{2}\left\| \underbrace{\sum_{i=1}^{K_1} 2\lambda_{1i}\, x_i^{(0)}} - \underbrace{\sum_{j} 2\lambda_{2j}\, x_j^{(1)}} \right\| \le \lambda_3$$

in conv hull of pts with label 1

$$\sum_{i=1}^{K_1} 2\lambda_{1i} = \sum_{j} 2\lambda_{2j} = \frac{1}{2} \times 2$$

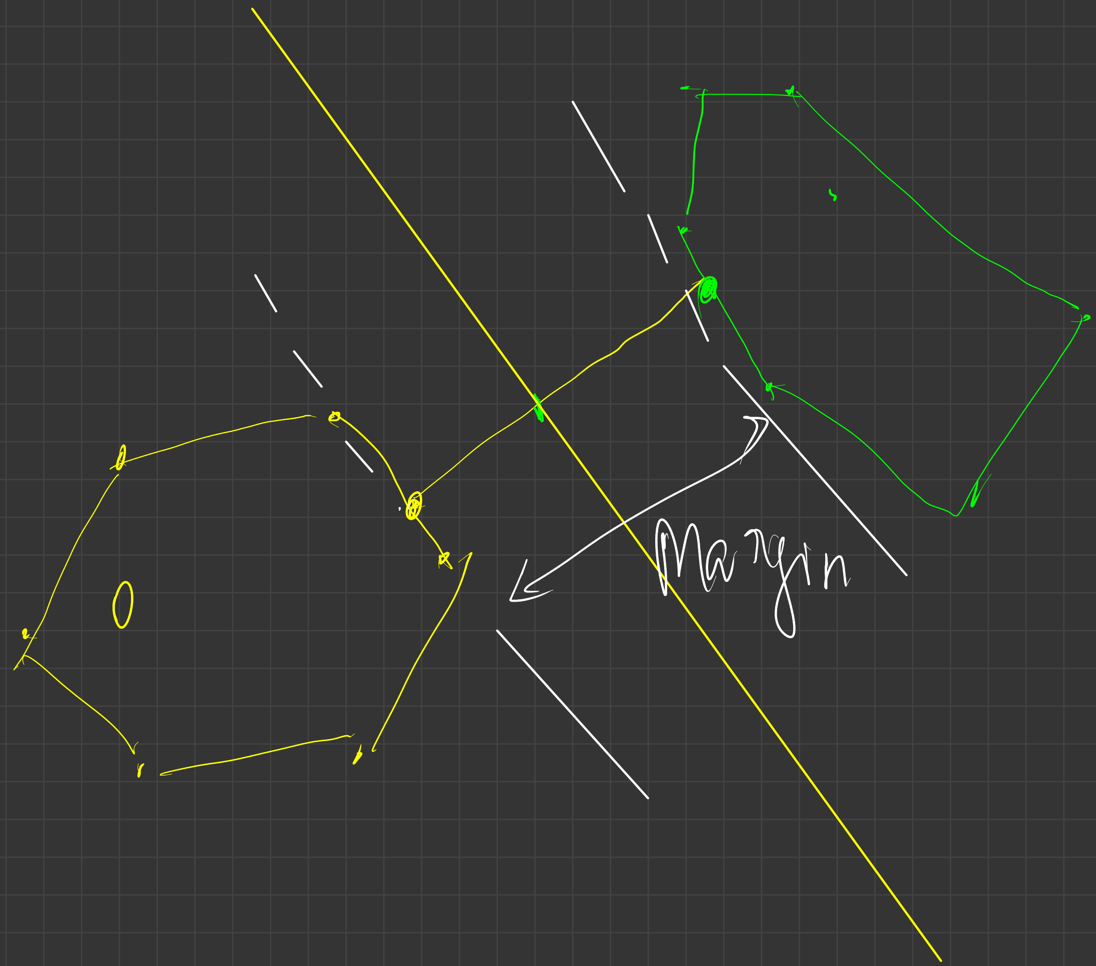
$$\underline{\lambda_1} \geq 0 \qquad \underline{\lambda_2} \geq 0 \qquad \lambda_3 \geq 0$$

Minimize $\frac{1}{2} \| \overbrace{\sum_{i=1}^{k_1} 2\lambda_{1i} x_i^{(0)}} - \overbrace{\sum_{j} 2\lambda_{2j} x_j^{(1)}} \|$

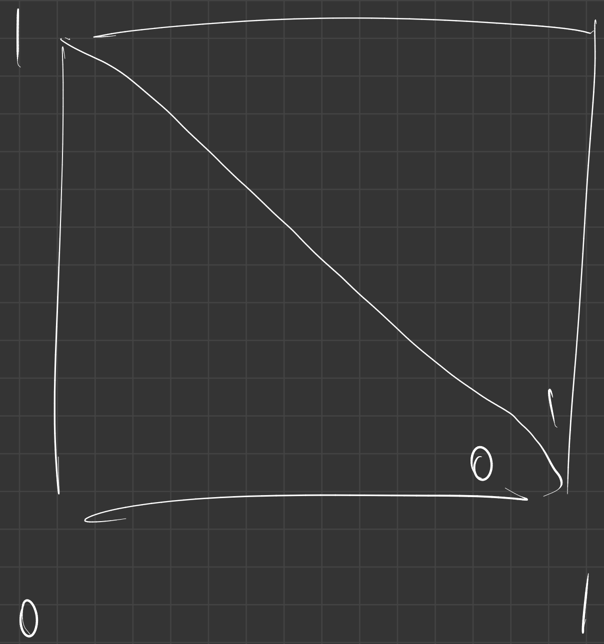
st $\sum_{i=1}^{k_1} 2\lambda_{1i} = \sum_{j} 2\lambda_{2j} = \frac{1}{2} \times 2$

$\lambda_1 \geq 0 \quad \lambda_2 \geq 0 \quad \lambda_3 \geq 0$

convex combination



Max margin = Min dist b/w convex hulls

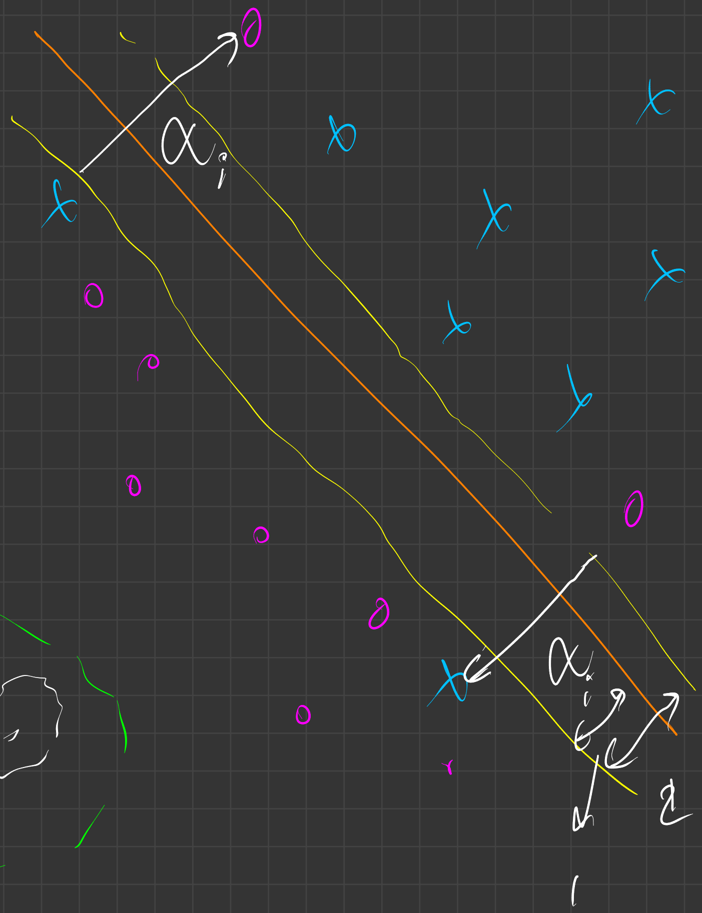

1
0
1
0
1

What if points are NOT linearly separable?

$y_i = \{-1, 1\}$ → new labels

For all points i,

$$\underbrace{y_i\left(a^T x_i - b\right) \ge 1}\ - \alpha_i$$



Constraints

t

$$y_i\left(\frac{a^T x_i}{\|a\|} - b\right) \ge \frac{1}{\|a\|} - \frac{\alpha_i}{\|a\|}$$

$ax \le b \Rightarrow -1$

$ax > b \Rightarrow +1$

Goal:

$$\text{Minimize} \quad \|\underline{a}\|^2 + \sum_{j=1}^{m} \alpha_j^p$$

ST

$$y_i \left(\underline{a}^T \underline{x}_i - b \right) \geq 1 - \alpha_i \quad \forall i$$

HW: Simulate this!

PRINCIPAL COMPONENT ANALYSIS

Given points sampled from a distribution p_x, along what directions is the "variation" the largest?

$$\underbrace{\underline{X}_1,\ \underline{X}_2,\ \cdots\ \underline{X}_N}_{\text{zero mean}}$$

Find unit norm $\underline{v}$ s.t. $\mathbb{E}(\underline{v}^T \underline{X})^2$ is maximized

Maximize $\mathbb{E}\left[(\underline{v}^T \underline{X})^2\right]$

ST $\|\underline{v}\| = 1$

Maximizing $\underline{v}$ is called the first principal component

Maximize $\mathbb{E}\left[(\underline{v}^T \underline{X})(X^T v)\right] = \mathbb{E}\left[v^T (X X^T) v\right]$

$= v^T (\mathbb{E} X X^T) v$

$= v^T \Sigma v$

covariance matrix

Maximize $v^T \Sigma v$ $\rightarrow$ symmetric PSD

s.t. $\|v\| = 1$

$\shortparallel$

Largest eigenvalue of Σ

$\underline{v}^* \rightarrow$ largest eigenvector.

k-principal components

Maximize $\mathbb{E} \| V^T \underline{x} \|$ $\rightarrow$ random vector $\mathbb{R}^n$

V: $n \times k$ matrix with orthonormal cols.

Maximizing $v^T \Sigma v$

if $\|v\|=1$

$\equiv$ Minimizing $\|vv^T - \Sigma\|_F^2$

$$\|vv^T - \Sigma\|_F^2 = tr\left((vv^T - \Sigma)^T (vv^T - \Sigma)\right)$$

$$= tr\left[vv^Tvv^T - 2vv^T\Sigma + \Sigma^T\Sigma\right]$$

$$= tr\left[(v^Tv)^2 + \Sigma^2 - 2vv^T\Sigma\right]$$

$$= 1 + tr(\Sigma^2) - 2v^T\Sigma v$$

Minimize $\| vv^T - \Sigma \|_F$

s.t. $\|v\| = 1$

$V = vv^T$

- rank $= 1$
- PSD
- $V^2 = V$ (projection matrix)

Assume variance$(X) = 1$

$\lambda_{max}(\Sigma) = 1$ (Prove this)

$$I - \Sigma \succeq 0$$

$$\text{Minimize} \quad \| V - \Sigma \|_F \longrightarrow \operatorname{tr}(\Sigma^T \Sigma V)$$

$$\text{St} \quad \operatorname{rank}(V) = 1 \longrightarrow \operatorname{tr}(V) = 1$$

$$V^2 - V = 0 \longrightarrow \text{relax this}$$

$$V \succeq 0$$

$$I - V \succeq 0$$

If V is a projection matrix, then eigenvalues are 0 or 1

Rank = Multiplicity of 1 = sum of eigenvalues = trace