



Lagrange Multipliers & Duality of SVMs

Rishabh Iyer

University of Texas at Dallas

Announcements (09/12/2022)

- Assignment 1 will be out by EOW
- Either next Mon or Wed, we will have a discussion on project ideas
- Midterm exam will be in the 2nd or 3rd week of October.

The Strategy So Far...

- Choose hypothesis space

Regression

$$y \in \mathbb{R}, f(x) = w^T x + b$$

Perception/SVM

$$y \in \{-1, +1\}$$

$$f(x) = \text{Sign}(w^T x + b)$$

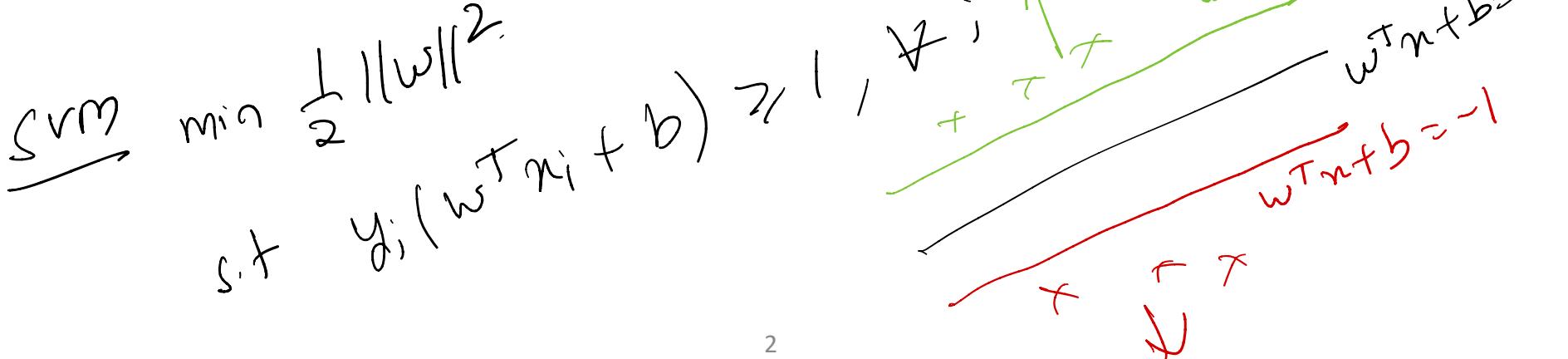
- Construct loss function (ideally convex)

$$L(y, f(x, w, b)) = [y - f(x, w, b)]^2 / L(y, f(x, w, b))$$

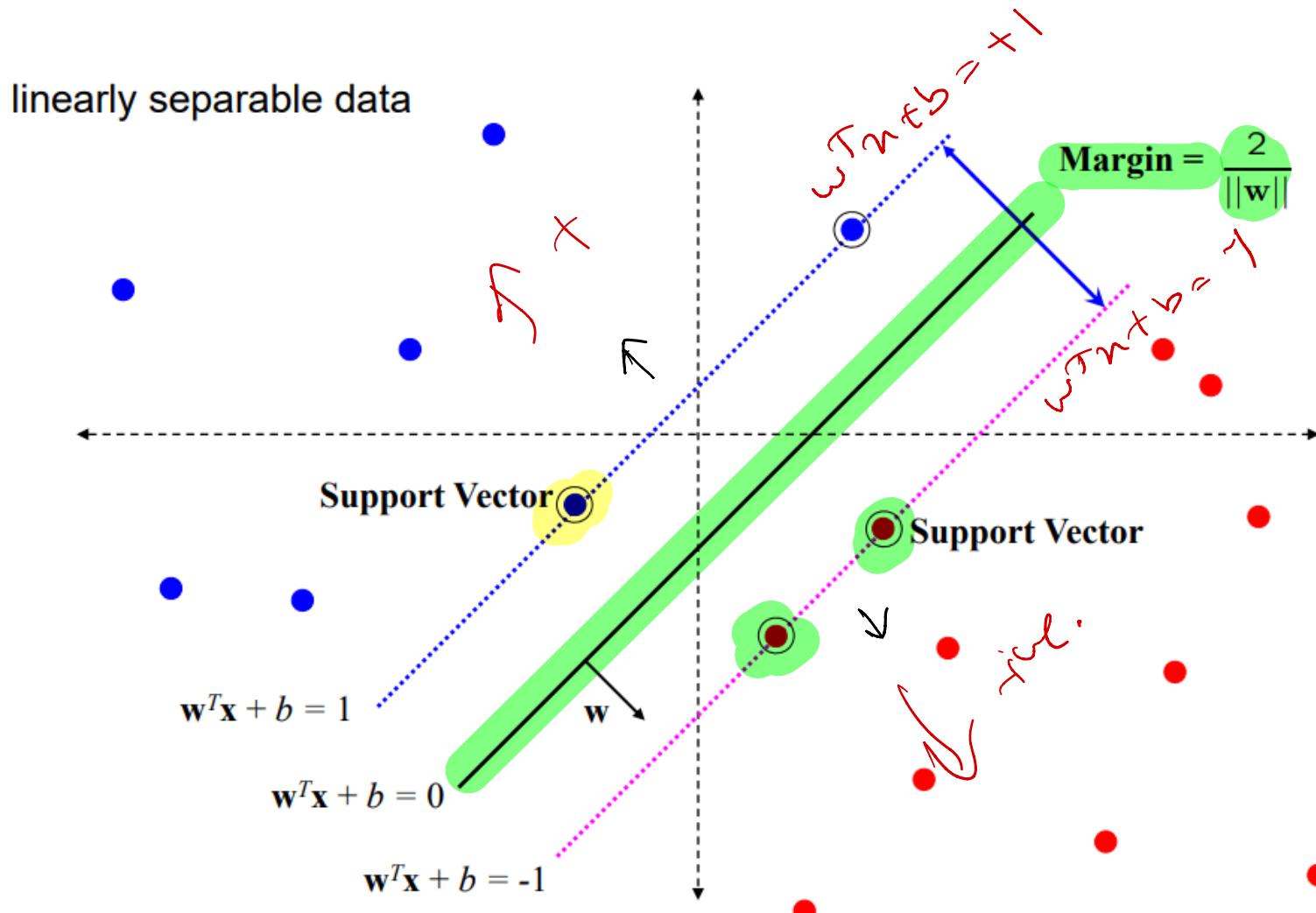
- Minimize loss to "learn" correct parameters

$$\sum_{i=1}^M L(y_i, f(x_i, w, b))$$

$$= \max(0) - y f(x, w, b)$$



Recap: SVM



SVM Optimization Problem

- Recall: The SVM optimization problem:

$$\underbrace{\min_{w,b} \|w\|^2}$$

such that

$$\underbrace{y^{(i)}}_{\textcolor{red}{-}} \left(\underbrace{w^T x^{(i)}}_{\textcolor{red}{-}} + b \right) \geq 1, \text{ for all } i$$

- This is a standard quadratic programming problem
 - Falls into the class of **convex optimization problems**
 - Can be solved with many specialized optimization tools (e.g., `quadprog()` in MATLAB)

Constrained Optimization



A mathematical detour, we'll come back to SVMs soon!

$$\min_{x \in \mathbb{R}^n} f_0(x)$$

subject to:

$x \rightarrow \text{opt variable.}$

$$x = \begin{bmatrix} w \\ b \end{bmatrix}$$

$$f_i(x) \leq 0, \quad i = 1, \dots, m$$

$$h_i(x) = 0, \quad i = 1, \dots, p$$

$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 \\ \text{s.t. } & y_i (w^T x_i + b) \geq 1 \\ & \Rightarrow 1 - y_i (w^T x_i + b) \leq 0, \quad i=1-\mathcal{M} \\ & f_i(w, b) \leq 0 \end{aligned}$$

Constrained Optimization

$$\min_{x \in \mathbb{R}^n} f_0(x)$$

f_0 is not necessarily convex

subject to:

$$f_i(x) \leq 0, \quad i = 1, \dots, m$$

$$h_i(x) = 0, \quad i = 1, \dots, p$$

General Optimization

$$\min_{x \in \mathbb{R}^n} f_0(x)$$

subject to:

$$\begin{aligned} f_i(x) &\leq 0, \\ h_i(x) &= 0, \end{aligned}$$

$$\begin{aligned} i &= 1, \dots, m \\ i &= 1, \dots, p \end{aligned}$$

Constraints do not need to
be linear

Example

$$\min_{x \in \mathbb{R}^2} f_0(x) = x_1 \log x_1 + x_2 \log x_2$$

$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

subject to:

$$x_1 + x_2 = 1 \quad \leftarrow h(x) = 0$$

$$x_1 \geq 0$$

$$x_2 \geq 0$$

$$x_1 + x_2 - 1 = 0$$

$$f_1(x) \leq 0$$

$$-x_1 \leq 0$$

$$f_2(x) \leq 0$$

$$-x_2 \leq 0$$

Example

$$\min_{x \in \mathbb{R}^2} x_1 \log x_1 + x_2 \log x_2$$

subject to:

$$\begin{aligned}1 - x_1 - x_2 &= 0 \\-x_1 &\leq 0 \\-x_2 &\leq 0\end{aligned}$$

Lagrangian

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

Constraints : $\lambda \geq 0$

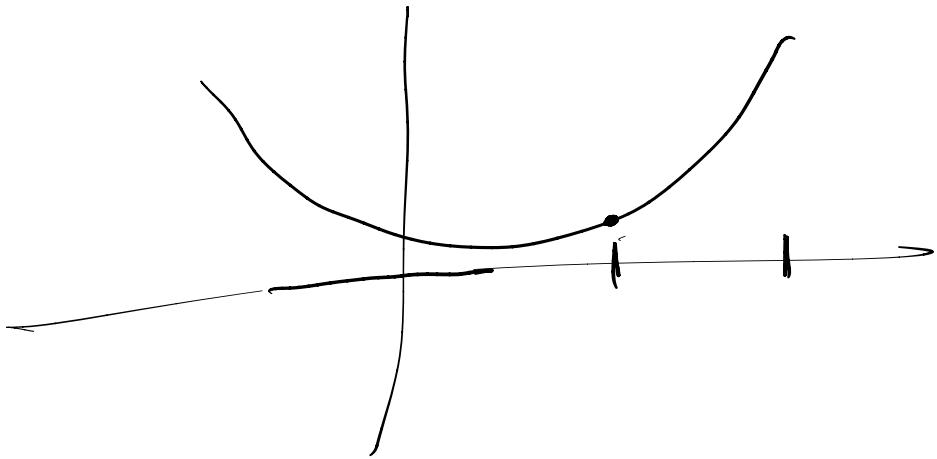
- Incorporate constraints into a new objective function
- $\lambda \geq 0$ and ν are vectors of **Lagrange multipliers**
- The Lagrange multipliers can be thought of as enforcing soft constraints

$$\begin{aligned} & \min f_0(\mathbf{x}) \\ \text{s.t. } & f_i(\mathbf{x}) \leq 0 \\ & h_i(\mathbf{x}) = 0 \end{aligned}$$

λ_i [$i = 1:m$]

ν_i [$i = 1:p$]

Why one cannot do GD when there are
Constraints?



Example

$$\min_{x \in \mathbb{R}^3} x_1 \log x_1 + x_2 \log x_2 = f_0(x)$$

subject to:

$$1 - x_1 - x_2 = 0$$

$$-x_1 \leq 0$$

$$-x_2 \leq 0$$

$$h(x) \leftarrow v_1$$

$$f_1(x) \leftarrow \gamma_1$$

$$f_2(x) \leftarrow \gamma_2$$

$$L(x_1, x_2, v_1, \lambda_1, \lambda_2)$$

$$= \underbrace{x_1 \log x_1 + x_2 \log x_2}_{f_0(x)} + v_1 \cdot \underbrace{(1 - x_1 - x_2)}_{h(x)} - \lambda_1 \underbrace{x_1}_{f_1(x)} - \underbrace{\lambda_2 x_2}_{f_2(x)}$$

Duality

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m f_i(x) + \sum_{r=1}^k \nu_r h_r(x)$$

- Construct a **dual function** by minimizing the Lagrangian over the primal variables

$$\min_{\lambda, \nu} \quad g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$$

- $g(\lambda, \nu) = -\infty$ whenever the Lagrangian is not bounded from below for a fixed λ and ν

$$\min f_0(x)$$

$$\text{s.t. } h_i(x) = 0, i=1:m$$

$$f_i(x) \leq 0, i=1:p$$

Primal

$$\max_{\lambda, \nu} g(\lambda, \nu)$$

$$\text{s.t. } \lambda \geq 0$$

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$$

Dual

Example

$$\min_{x \in \mathbb{R}^2} \underbrace{x_1 \log x_1 + x_2 \log x_2}_{f_0(x)} \quad \text{↗}$$

subject to:

$$1 - x_1 - x_2 = 0 \quad \leftarrow v_1$$

$$-x_1 \leq 0 \quad \leftarrow \gamma_1$$

$$-x_2 \leq 0 \quad \leftarrow \gamma_2$$

$$L(x, \gamma, v) = x_1 \log x_1 + x_2 \log x_2 + v_1(1 - x_1 - x_2) - \gamma_1 x_1 - \gamma_2 x_2$$

$$g(\underline{\gamma}, \underline{v}) = \min_x L(x, \gamma, v)$$

$$\frac{dL}{dx_1} = \log x_1 + 1 - v_1 - \gamma_1 = 0 \Rightarrow x_1 = \exp(v_1 + \gamma_1 - 1)$$

$$\frac{dL}{dx_2} = \log x_2 + 1 - v_1 - \gamma_2 = 0 \Rightarrow x_2 = \exp(v_1 + \gamma_2 - 1)$$

$$\frac{dL}{dx_2} = \log x_2 + 1 - v_1 - \gamma_2 = 0 \Rightarrow x_2 = \exp(v_1 + \gamma_2 - 1)$$

Example

$$\min_{x \in \mathbb{R}^3} x_1 \log x_1 + x_2 \log x_2$$

$$\max g(\gamma, v)$$

subject to:

$$1 - x_1 - x_2 = 0$$

$$-x_1 \leq 0$$

$$-x_2 \leq 0$$

$$\begin{array}{l} s.t. \\ \gamma \geq 0 \end{array}$$

Dual.

$$g(\gamma, v) = (v_1 + \gamma_1 - 1) \underbrace{\exp(v_1 + \gamma_1 - 1)}_{\gamma_1 \exp(v_1 + \gamma_1 - 1)} + (v_2 + \gamma_2 - 1) \exp(v_2 + \gamma_2 - 1)$$

$$+ v_1 (1 - \exp(v_1 + \gamma_1 - 1) - \exp(v_2 + \gamma_2 - 1))$$

$$- \gamma_1 \exp(v_1 + \gamma_1 - 1) - \gamma_2 (\exp(v_2 + \gamma_2 - 1))$$

$$= v_1 - \exp(v_1 + \gamma_1 - 1) - \exp(v_2 + \gamma_2 - 1)$$

$$\max (v_1 - \exp(v_1 + \gamma_1 - 1) - \exp(v_2 + \gamma_2 - 1))$$

① Write down the constrained formulation

$$\min f_0(x)$$

$$\text{s.t. } f_i(x) \leq 0, i = 1:m$$

$$h_i(x) = 0, i = 1:P$$

② $L(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^P v_i h_i(x)$

③ $g(v, \lambda) = \min_x \left[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^P v_i h_i(x) \right]$

$$\frac{\partial L}{\partial x} = \underbrace{\frac{\partial f_0}{\partial x}}_{\lambda = F(\lambda, v)} + \sum_{i=1}^m \lambda_i \frac{\partial f_i}{\partial x} + \sum_{i=1}^P v_i \frac{\partial h_i}{\partial x} = 0$$

$\lambda = F(\lambda, v)$

$$\begin{array}{ll} \min_{w,b} & \frac{1}{2} \|w\|^2 \\ \text{s.t.} & y_i(w^T n_i + b) \geq 1 \\ & 1 - y_i(w^T n_i + b) \leq 0 \\ & f_i(w, b) \leq 0 \end{array}$$

Primal.

↓ Lagrangian.

$$L(w, b, \gamma) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^M \gamma_i (1 - y_i(w^T n_i + b))$$

$$g(\gamma) = \min_{w,b} \left[\frac{1}{2} \|w\|^2 + \sum_{i=1}^M \gamma_i (1 - y_i(w^T n_i + b)) \right]$$

$$\frac{\partial L}{\partial w} = 0, \frac{\partial L}{\partial b} = 0$$

$$w = F(\gamma), b = F(\gamma)$$

The Primal Problem

$$\min_{x \in \mathbb{R}^n} f_0(x)$$

subject to:

$$\begin{aligned} f_i(x) &\leq 0, & i = 1, \dots, m \\ h_i(x) &= 0, & i = 1, \dots, p \end{aligned}$$

Equivalently,

$$\min_x \max_{\lambda \geq 0, \nu} L(x, \lambda, \nu)$$

$$f_0(n)$$

Why are these equivalent?

The Primal Problem

$$\min_{x \in \mathbb{R}^n} f_0(x)$$

subject to:

$$\begin{aligned} f_i(x) &\leq 0, & i = 1, \dots, m \\ h_i(x) &= 0, & i = 1, \dots, p \end{aligned}$$

Equivalently,

$$\inf_x \sup_{\lambda \geq 0, \nu} L(x, \lambda, \nu) \leq 0$$

$$\sup_{\lambda \geq 0, \nu} \left[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right] = \infty$$

f_i(x) ≤ 0
 h_i(x) = 0

whenever x violates the constraints

The Dual Problem

$$\underset{\lambda \geq 0, \nu}{\sup} \underset{\text{Concave}}{\underbrace{g(\lambda, \nu)}}$$

(max)

Equivalently,

$$\underset{\lambda \geq 0, \nu}{\sup} \underset{x}{\inf} L(x, \lambda, \nu)$$

↑
Concave

- The dual problem is always concave, even if the primal problem is not convex
 - For each x , $L(x, \lambda, \nu)$ is a linear function in λ and ν
 - Minimum (or infimum) of linear functions is concave!

Convexity

$$f(x) = \underset{y}{\inf} (\sup g(x, y))$$

Concavity

$$f(x) = \underset{y}{\min} (\inf g(x, y))$$

$$\min_x f_0(x)$$

$$\text{s.t. } f_i(x) \leq 0, i=1 \dots m$$

→ P

$$h_i(x) = 0, i=1 \dots p$$

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p v_i h_i(x)$$

$$(P) \equiv \min_x \underbrace{\max_{\lambda \geq 0, v} L(x, \lambda, v)}_{f_0(x)}$$

$$(D) \equiv \max_{\lambda \geq 0, v} \min_x L(x, \lambda, v)$$

Primal vs. Dual



- Why?

① $g(\lambda, \nu) \leq L(x, \lambda, \nu)$ for all x

v) • $L(x', \lambda, \nu) \leq f_0(x')$ for any feasible $x', \lambda \geq 0$

- x is **feasible** if it satisfies all of the constraints

- Let x^* be the optimal solution to the primal problem and $\lambda \geq 0$

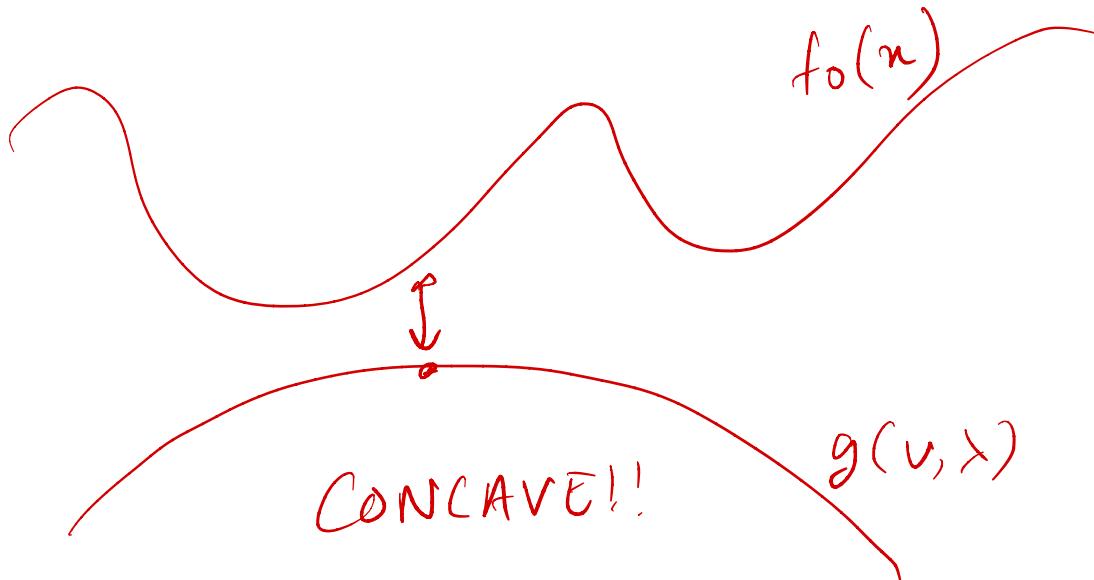
$$\geq 0$$

Dual. \rightarrow $\underbrace{g(\lambda, \nu)}_{\text{Dual}} \leq \underbrace{L(x^*, \lambda, \nu)}_{\text{Optimality Condition}} \leq f_0(x^*)$ \rightarrow Primal

$$g(n^*, r^*) \leq f_0(n^*)$$

$$f^*(x^*) \leq f_0(x^*)$$

$$L(x, \gamma, v) = f_0(n) + \sum_{i=1}^m \underbrace{f_i(n)}_{\leq 0} + \sum_{i=1}^p \underbrace{v_i m^{(n)}}_{\leq 0}$$



$$\min f_0(x)$$

$$\max g(\gamma v)$$

$$\text{s.t. } f_i(x) \leq 0, i=1:M$$

$$\gamma \geq 0$$

$$n_i(x) \geq 0, i=1:P$$

$$L(x, \gamma v) = f_0(x) + \sum_{i=1}^M f_i(x) + \sum_{i=1}^P n_i(x)$$

$$g(\gamma) = \max_{x \in \mathcal{X}} L(x, \gamma v)$$

$$x = P(\gamma v)$$

Example: Solving the Dual Problem



$$\min_{x \in \mathbb{R}^3} x_1 \log x_1 + x_2 \log x_2$$

subject to:

$$\begin{aligned} 1 - x_1 - x_2 &= 0 \\ -x_1 &\leq 0 \\ -x_2 &\leq 0 \end{aligned}$$

$$\begin{aligned} \max_{\lambda, v} g(\lambda, v) \\ \text{s.t. } & \quad \lambda \geq 0, v \geq 0 \end{aligned}$$

$$\begin{aligned} \max_{\lambda, v} g(\lambda, v) &= \max_{\lambda, v} v_1 - \underbrace{\exp(v_1 + \lambda_1 - 1)}_{\text{dec}(\lambda_1)} - \underbrace{\exp(v_1 + \lambda_2 - 1)}_{\text{dec}(\lambda_2)} \\ &\quad \lambda \geq 0 \\ &\quad \text{max attained at } (\lambda_1, \lambda_2) = (0, 0) \\ \max_{v_1} v_1 - 2\exp(v_1 - 1) &= g(\lambda, v) \\ \frac{dv}{dv_1} &= 0 \Rightarrow 1 - 2\exp(v_1 - 1) = 0 \\ \Rightarrow v_1 &= \log(0.5) + 1 \end{aligned}$$

More Examples

- Minimize $x^2 + y^2$ subject to $x + y \geq 1$
- Given a point $z \in \mathbb{R}^n$ and a hyperplane $w^T x + b = 0$, find the projection of the point z onto the hyperplane

$$\begin{aligned} & \min_x \|x - z\|^2 \\ & \text{s.t. } w^T x + b = 0 \end{aligned}$$

Duality

- Under certain conditions, the two optimization problems are equivalent

$$\sup_{\lambda \geq 0, \nu} \inf_x L(x, \lambda, \nu) = \inf_x \sup_{\lambda \geq 0, \nu} L(x, \lambda, \nu)$$

- This is called **strong duality**.

- If the inequality is strict, then we say that there is a **duality gap**
 - Size of gap measured by the difference between the two sides of the inequality

Slater's Condition

For any optimization problem of the form

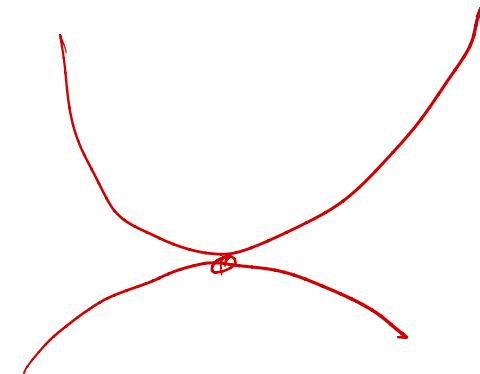
$$\min_{x \in \mathbb{R}^n} f_0(x)$$

subject to:

$$\begin{aligned} f_i(x) &\leq 0, \quad i = 1, \dots, m \\ Ax &= b \end{aligned}$$

where f_0, \dots, f_m are **convex functions**, strong duality holds if there exists an x such that

$$\begin{aligned} f_i(x) &< 0, \quad i = 1, \dots, m \\ Ax &= b \end{aligned}$$



Dual SVM

$$\begin{aligned} w &\in \mathbb{R}^n \\ b &\in \mathbb{R}^M \end{aligned}$$

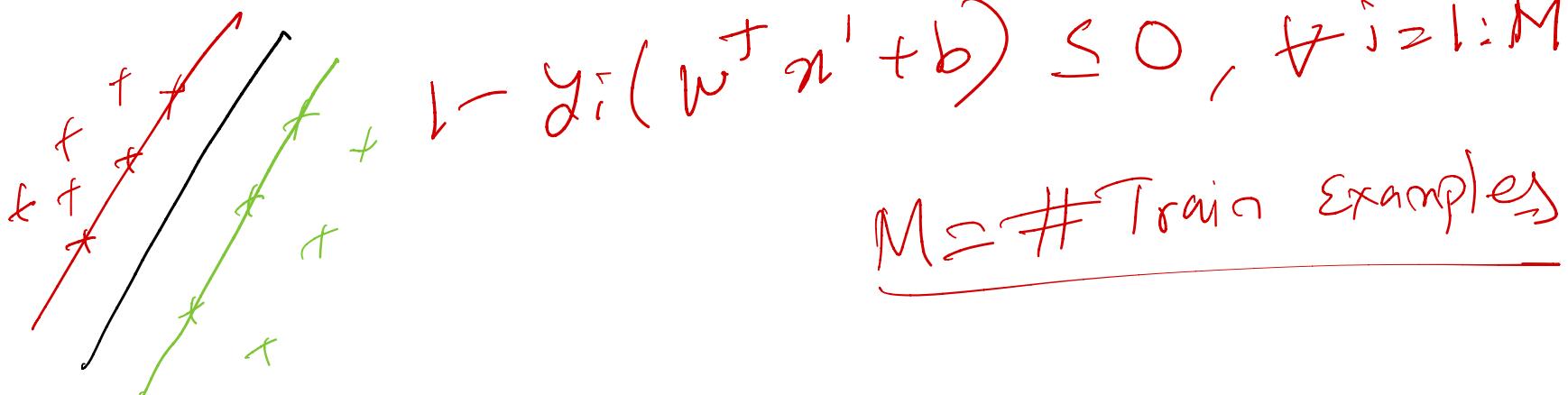
such that

$$\min_w \frac{1}{2} \|w\|^2 \quad \begin{matrix} \leftarrow \text{Convex} & \checkmark \end{matrix}$$

$$y_i(w^T x^{(i)} + b) \geq 1, \text{ for all } i$$

$$\begin{aligned} f_i(w, b) &\leq 0 \\ 1 - y_i(w^T x^{(i)} + b) &\leq 0 \end{aligned}$$

- Note that Slater's condition holds as long as the data is linearly separable



Dual SVM



$$L(w, b, \lambda) = \frac{1}{2} w^T w + \sum_i \lambda_i (1 - y_i (w^T x^{(i)} + b))$$

$\lambda \in \mathbb{R}^m$

Convex in w , so take derivatives to form the dual

$$\frac{\partial L}{\partial w_k} = w_k + \sum_i -\lambda_i y_i x_k^{(i)} = 0$$

$\lambda \geq 0$

$$\boxed{\frac{\partial L}{\partial b} = \sum_i -\lambda_i y_i = 0} \Rightarrow \sum_i \lambda_i y_i = 0$$

$$w = \sum_{i=1}^m \lambda_i y_i x^{(i)}$$

Dual SVM

$$L(w, b, \lambda) = \frac{1}{2} w^T w + \sum_i \lambda_i (1 - y_i (w^T x^{(i)} + b))$$

Convex in w , so take derivatives to form the dual

$$\bar{w} = \sum_i \lambda_i y_i x^{(i)}$$
$$\sum_i \lambda_i y_i = 0$$

constraint

Dual SVM

$$\max_{\lambda \geq 0} -\frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j x^{(i)T} x^{(j)} + \sum_i \lambda_i$$

↑↑↑
such that
 $\sum_i \lambda_i y_i = 0$

Convex Maximization
 Convex Minimization

$g(\gamma)$

- By strong duality, solving this problem is equivalent to solving the primal problem
 - Given the optimal λ , we can easily construct w (b can be found by **complementary slackness...**)

$$\frac{1}{2} \|\omega\|^2 + \sum_{i=1}^m z_i (1 - y_i (\omega^\top x^{(i)} + b))$$

$$\|\omega\| = \sqrt{\sum_{i=1}^m z_i y_i x^{(i)T} \omega}$$

$$\Rightarrow \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m z_i z_j y_i y_j x^{(i)T} x^{(j)}$$

$$+ \sum_{i=1}^m z_i (1 - y_i \sum_{j=1}^m z_j y_j x^{(j)T} \omega)$$

$$- y_i b$$

mat

$$\left[-\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m z_i z_j y_i y_j x^{(i)T} x^{(j)} \right]$$

$$+ \sum_{i=1}^m z_i$$

$$s.t \quad z \geq 0$$

$$\sum_i z_i y_i = 0$$

Primal

$$\min_{w, b} \frac{1}{2} \|w\|^2$$

$$\text{subject to } y_i(w^T x^{(i)} - b) \geq 1$$

Equivalent

Dual.

max



$$-\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \gamma_i \gamma_j y_i y_j x^{(i)^T} x^{(j)}$$

s.t.

$$\gamma \geq 0$$

$$\sum_i \gamma_i y_i = 0$$



constraints

$$w = \sum_{i=1}^m \gamma_i y_i x^{(i)}$$

Complementary Slackness

- Suppose that there is zero duality gap
 (ω^*, b^*)
- Let x^* be an optimum of the primal and (λ^*, ν^*) be an optimum of the dual

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

$$= \inf_x \left[f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right]$$

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

$$= f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*)$$

$$\leq f_0(x^*)$$

$$\max g(\zeta, v)$$

$$\geq 0$$

$$\min f_0(x)$$

$$+$$

$$\text{s.t } h_i(x) = 0$$

$$f_i(x) \leq 0$$

$$h_i(x^*) = 0$$

$$f_i(x^*) \leq 0$$

$$x^* \geq 0$$

Complementary Slackness

- This means that

$$\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0$$

- As $\lambda \geq 0$ and $f_i(x_i^*) \leq 0$, this can only happen if $\lambda_i^* f_i(x^*) = 0$ for all i
- Put another way,
 - If $f_i(x^*) < 0$ (i.e., the constraint is not tight), then $\lambda_i^* = 0$
 - If $\lambda_i^* > 0$, then $f_i(x^*) = 0$

Dual SVM (Obtaining b)

$$\max_{\lambda \geq 0} -\frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j x^{(i)T} x^{(j)} + \sum_i \lambda_i$$

$y_i = \pm 1$

such that

$$\sum_i \lambda_i y_i = 0$$

$$f_i(w^*, b^*) = 0$$

- By complementary slackness, $\lambda_i^* > 0$ means that $x^{(i)}$ is a support vector (can then solve for b using w)
- In particular,

$$b = y_i - \underbrace{w \cdot x_i}_{}$$

for any i where $\lambda_i > 0$

$$b = \underbrace{\lambda_i}_{y_i} - w^T x_i$$

Dual SVM

$$\max_{\lambda \geq 0} -\frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \lambda_i \lambda_j y_i y_j \underbrace{x^{(i)T} x^{(j)}}_{M \times M} + \sum_i \lambda_i$$

M # training

such that

$$\sum_i \lambda_i y_i = 0$$

$O(M^2)$

- Takes $O(M^2)$ time just to evaluate the objective function
 - Active area of research to try to speed this up

$M \approx 10M.$

$M \times M$ 10^{14}