ECE595 / STAT598: Machine Learning I
Lecture 20.2: Support Vector Machine - Dual SVM

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

Outline

Support Vector Machine

- Lecture 19 SVM 1: The Concept of Max-Margin
- Lecture 20 SVM 2: Dual SVM
- Lecture 21 SVM 3: Kernel SVM

This lecture: Support Vector Machine: Duality

- Lagrange Duality
  - Maximize the dual variable
  - Minimax Problem
  - Toy Example
- Dual SVM
  - Formulation
  - Interpretation

Dual of SVM

- We want to find the dual problem of

$$\begin{aligned} \underset{\boldsymbol{w}, w_0}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 \\ \text{subject to} \quad & y_j(\boldsymbol{w}^T\boldsymbol{x}_j + w_0) \geq 1, \quad j = 1, \ldots, N. \end{aligned}$$

- We start with the Lagrangian function

$$\mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\lambda}) \stackrel{\text{def}}{=} \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \sum_{j=1}^{N} \lambda_j \left[1 - y_j(\boldsymbol{w}^T\boldsymbol{x}_j + w_0)\right].$$

- Let us **minimize** over $(\boldsymbol{w}, w_0)$:

$$\nabla_{\boldsymbol{w}}\mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\lambda}) = \boldsymbol{w} - \sum_{j=1}^{N} \lambda_j y_j \boldsymbol{x}_j = \boldsymbol{0}$$

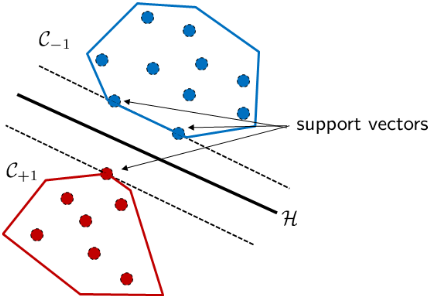
$$\nabla_{w_0}\mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\lambda}) = \sum_{j=1}^{N} \lambda_j y_j = 0.$$

Interpreting $\nabla_{\boldsymbol{w}}\mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\lambda}) = 0$

- The first result suggests that

$$\boldsymbol{w} = \sum_{j=1}^{N} \lambda_j y_j \boldsymbol{x}_j.$$

- This is support vector: λ_j is either $\lambda_j = 0$ or $\lambda_j > 0$.



Interpreting $\nabla_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\lambda}) = 0$

- The complementarity condition states that

$$\lambda_j^* \left[1 - y_j(\boldsymbol{w}^{*T} \boldsymbol{x}_j + w_0^*) \right] = 0, \quad \text{for} \quad j = 1, \ldots, N.$$

- If $1 - y_j(\boldsymbol{w}^{*T} \boldsymbol{x}_j + w_0^*) > 0$, then $\lambda_j^* = 0$
- If $\lambda_j^* > 0$, then $1 - y_j(\boldsymbol{w}^{*T} \boldsymbol{x}_j + w_0^*) = 0$
- So you can define the **support vector set**:

$$\mathcal{V} \overset{\text{def}}{=} \{ j \mid \lambda_j^* > 0 \}.$$

- So the optimal weight is

$$\boldsymbol{w}^* = \sum_{j \in \mathcal{V}} \lambda_j^* y_j \boldsymbol{x}_j.$$

Back to Duality

The Lagrangian function is

$$\mathcal{L}(\boldsymbol{w}^*, w_0^*, \boldsymbol{\lambda}) = \frac{1}{2}\|\boldsymbol{w}^*\|_2^2 + \sum_{j=1}^{N} \lambda_j \left[1 - y_j((\boldsymbol{w}^*)^T \boldsymbol{x}_j + w_0)\right]$$

$$= \underbrace{\frac{1}{2}\left\| \sum_{j=1}^{N} \lambda_j y_j \boldsymbol{x}_j \right\|_2^2}_{A}$$

$$+ \underbrace{\sum_{j=1}^{N} \lambda_j \left[1 - y_j \left(\left(\sum_{i=1}^{n} \lambda_i y_i \boldsymbol{x}_i \right)^T \boldsymbol{x}_j + w_0 \right)\right]}_{B}$$

Back to Duality

- We can show that

$$A = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j$$

$$B = \sum_{j=1}^{N}\lambda_j - \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j - \underbrace{\left(\sum_{j=1}^{N}\lambda_j y_j\right)}_{=0} w_0$$

- and we can show that

$$A + B = \sum_{j=1}^{N}\lambda_j + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j$$

Back to Duality

- Therefore, the dual problem is

$$\underset{\boldsymbol{\lambda} \geq 0}{\text{maximize}} \quad -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j + \sum_{j=1}^{N} \lambda_j$$

$$\text{subject to} \quad \sum_{j=1}^{N} \lambda_j y_j = 0.$$

- If you prefer matrix-vector:

$$\underset{\boldsymbol{\lambda} \geq 0}{\text{maximize}} \quad -\frac{1}{2} \boldsymbol{\lambda}^T \boldsymbol{Q} \boldsymbol{\lambda} + \boldsymbol{1}^T \boldsymbol{\lambda}$$

$$\text{subject to} \quad \boldsymbol{y}^T \boldsymbol{\lambda} = \boldsymbol{0}.$$

- We can combine the constraints $\boldsymbol{\lambda} \geq 0$ and $\boldsymbol{y}^T \boldsymbol{\lambda} = \boldsymbol{0}$ as

$$\boldsymbol{A} \boldsymbol{\lambda} \geq \boldsymbol{0}.$$

Back to Duality

- $\boldsymbol{y}^T\boldsymbol{\lambda} = 0$ means

$$\boldsymbol{y}^T\boldsymbol{\lambda} \geq 0, \quad \text{and} \quad \boldsymbol{y}^T\boldsymbol{\lambda} \leq 0.$$

- Thus, we can write $\boldsymbol{y}^T\boldsymbol{\lambda} = 0$ as

$$\begin{bmatrix} \boldsymbol{y}^T \\ -\boldsymbol{y}^T \end{bmatrix} \boldsymbol{\lambda} \geq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

- Therefore, the matrices $\boldsymbol{Q}$ and $\boldsymbol{A}$ are

$$\boldsymbol{Q} = \begin{bmatrix} y_1 y_1 \boldsymbol{x}_1^T\boldsymbol{x}_1 & \dots & y_1 y_N \boldsymbol{x}_1^T\boldsymbol{x}_N \\ y_2 y_1 \boldsymbol{x}_2^T\boldsymbol{x}_1 & \dots & y_2 y_N \boldsymbol{x}_2^T\boldsymbol{x}_N \\ \vdots & \vdots & \vdots \\ y_N y_1 \boldsymbol{x}_N^T\boldsymbol{x}_1 & \dots & y_N y_N \boldsymbol{x}_N^T\boldsymbol{x}_N \end{bmatrix} \quad \text{and} \quad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{y}^T \\ -\boldsymbol{y}^T \\ \boldsymbol{I} \end{bmatrix}$$

So How to Solve the SVM Problem?

- You look at the dual problem

$$\begin{aligned} \underset{\boldsymbol{\lambda}}{\text{maximize}} \quad & -\frac{1}{2}\boldsymbol{\lambda}^T \boldsymbol{Q}\boldsymbol{\lambda} + \mathbf{1}^T\boldsymbol{\lambda} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned}$$

- You get the solution $\boldsymbol{\lambda}^*$.
- Then compute $\boldsymbol{w}^*$:

$$\boldsymbol{w}^* = \sum_{j \in \mathcal{V}} \lambda_j^* y_j \boldsymbol{x}_j.$$

- $\mathcal{V}$ is the set of support vectors: $\lambda_j > 0$.

Are We Done Yet?

- Not quite.
- We still need to find out w_0^*.
- Pick any **support vector** $\boldsymbol{x}^+ \in \mathcal{C}_+$ and $\boldsymbol{x}^- \in \mathcal{C}_-$.
- Then we have

$$\boldsymbol{w}^T\boldsymbol{x}^+ + w_0 = +1, \quad \text{and} \quad \boldsymbol{w}^T\boldsymbol{x}^- + w_0 = -1.$$

- Sum them, we have $\boldsymbol{w}^T(\boldsymbol{x}^+ + \boldsymbol{x}^-) + 2w_0 = 0$, which means

$$w_0^* = -\frac{(\boldsymbol{x}^+ + \boldsymbol{x}^-)^T\boldsymbol{w}^*}{2}$$

Summary of Dual SVM

- Primal

$$\underset{\boldsymbol{w},w_0}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2$$
$$\text{subject to} \quad y_j(\boldsymbol{w}^T\boldsymbol{x}_j + w_0) \geq 1, \quad j = 1,\ldots,N.$$

- Strong Duality

$$\underbrace{\min_{\boldsymbol{w},w_0} \max_{\boldsymbol{\lambda}\geq 0} \mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\lambda})}_{\text{primal}} = \underbrace{\max_{\boldsymbol{\lambda}\geq 0} \min_{\boldsymbol{w},w_0} \mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\lambda})}_{\text{dual}}$$

- Dual

$$\underset{\boldsymbol{\lambda}\geq 0}{\text{maximize}} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j + \sum_{j=1}^{N}\lambda_j$$
$$\text{subject to} \quad \sum_{j=1}^{N}\lambda_j y_j = 0.$$

Summary of Dual SVM

- The weights are computed as

$$\boldsymbol{w}^* = \sum_{j=1}^{N} \lambda_j^* y_j \boldsymbol{x}_j.$$

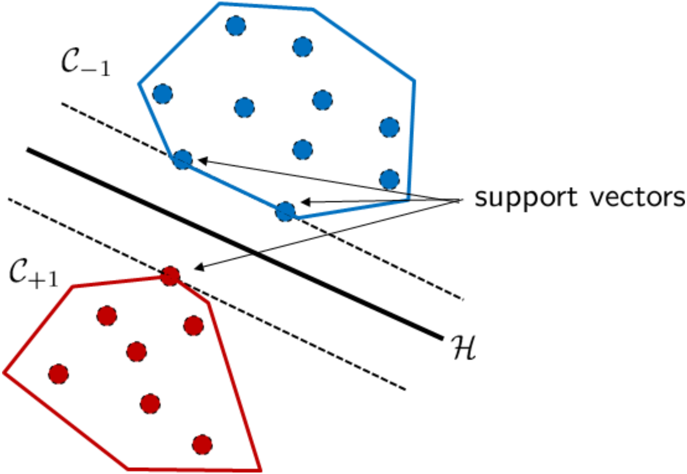
- This is support vector: λ_j is either $\lambda_j = 0$ or $\lambda_j > 0$.
- Pick any **support vector** $\boldsymbol{x}^+ \in \mathcal{C}_+$ and $\boldsymbol{x}^- \in \mathcal{C}_-$.
- Then we have

$$\boldsymbol{w}^T \boldsymbol{x}^+ + w_0 = +1, \quad \text{and} \quad \boldsymbol{w}^T \boldsymbol{x}^- + w_0 = -1.$$

- Sum them, we have $\boldsymbol{w}^T(\boldsymbol{x}^+ + \boldsymbol{x}^-) + 2w_0 = 0$, which means

$$w_0^* = -\frac{(\boldsymbol{x}^+ + \boldsymbol{x}^-)^T \boldsymbol{w}^*}{2}$$

Summary of Dual SVM



Support Vector Machine

- Mustafa, *Learning from Data*, e-Chapter
- Duda-Hart-Stork, *Pattern Classification*, Chapter 5.5
- Chris Bishop, *Pattern Recognition*, Chapter 7.1
- UCSD Statistical Learning
  http://www.svcl.ucsd.edu/courses/ece271B-F09/