CS 307 Lecture 14

Max Margin Classifier: Find a hyperplane that separates training set such that distance of closest point to hyperplane is maximized.

Hard SVM: When training set $S = \{(\underbrace{x^{(i)}}_{\in \mathbb{R}^d}, \underbrace{y^{(i)}}_{\in\{+1,-1\}})\}_{i=1}^n$ is linearly separable:

Classifier: $\theta^T x + \theta_0$, $\theta \to \in \mathbb{R}^d$

$$\min \quad \frac{1}{2}\|\theta\|_2^2 = \frac{1}{2}(\theta^T\theta)$$

$$\text{s.t.} \quad y^{(i)}(\theta^T x^{(i)} + \theta_0) \geq 1 \quad \forall i \in \{1 \ldots n\}$$

Soft SVM: When you want to allow the linear classifier to misclassify some examples in the training set.

$$\min \quad \frac{\overset{\text{hyperparameter}}{\lambda}}{2}\|\theta\|^2 + \frac{1}{n}\sum_{i=1}^n \underbrace{\max(0, 1 - y^{(i)}(\theta^T x^{(i)} + \theta_0))}_{\text{hinge loss}}$$

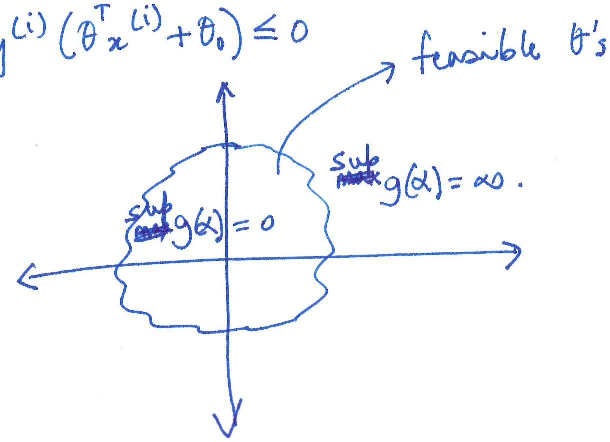
Hard SVM

$$\min \ \frac{1}{2}\|\theta\|^2 \qquad\qquad \min \ \frac{1}{2}\|\theta\|^2$$

$$\text{s.t } y^{(i)}(\theta^T x^{(i)} + \theta_0) \geq 1 \ \forall i \quad \longleftrightarrow \quad \text{s.t } \underbrace{1 - y^{(i)}(\theta^T x^{(i)} + \theta_0) \leq 0}_{n \text{ constraints}} \ \forall i$$

$$\underset{\in \mathbb{R}^n}{g(\alpha)} = \sum_{i=1}^n \alpha_i \left(1 - y^{(i)}(\theta^T x^{(i)} + \theta_0)\right) \qquad [\forall i \ \alpha_i \geq 0]$$

$$\underset{\alpha}{\overset{\text{sup}}{\cancel{\max}}} \ g(\alpha) = \begin{cases} 0 & \forall i \ 1 - y^{(i)}(\theta^T x^{(i)} + \theta_0) \leq 0 \\ \infty & \text{o.w.} \end{cases}$$



Lagrangian:

$$\mathcal{L}(\theta,\theta_0,\alpha) = \frac{1}{2}\|\theta\|^2 + \sum_{i=1}^{n} \alpha_i \left(1 - y^{(i)}(\theta^T x^{(i)} + \theta_0)\right)$$

$$\sup_{\alpha} \mathcal{L}(\theta,\theta_0,\alpha) = \begin{cases} \frac{1}{2}\|\theta\|^2 & \text{if } \forall i \; 1 - y^{(i)}(\theta^T x^{(i)}+\theta_0) \le 0 \\ \infty & \text{o.w.} \end{cases}$$

Hard SVM is equivalent to $\min_{\theta,\theta_0} \sup_{\alpha} \mathcal{L}(\theta,\theta_0,\alpha)$
"primal"

$$\sup_{\alpha} \min_{\theta,\theta_0} \mathcal{L}(\theta,\theta_0,\alpha) \le \min_{\theta,\theta_0} \sup_{\alpha} \mathcal{L}(\theta,\theta_0,\alpha) \quad \text{[weak duality]}$$

$=$ in our special case.

$$\min_{\theta,\theta_0} \mathcal{L}(\theta,\theta_0,\alpha) = \min_{\theta,\theta_0} \underbrace{\frac{1}{2}\|\theta\|^2 + \sum_{i=1}^{n} \alpha_i \left(1 - y^{(i)}(\theta^T x^{(i)} + \theta_0)\right)}_{} \quad \text{(convex)}$$

$$\nabla_{\theta} \mathcal{L}(\theta,\theta_0,\alpha) = \theta - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

When $\nabla_\theta \mathcal{L}(\theta,\theta_0,\alpha) = 0$,

$$\theta = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}. \leftarrow$$

$(A+B)^T = A^T + B^T$

$(a+b)(c+d) = ac + bc + ad + bd$

$$\mathcal{L}\left(\sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}, \theta_0, \alpha\right) = \frac{1}{2}\left(\sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}\right)^T \left(\sum_{j=1}^{n} \alpha_j y^{(j)} x^{(j)}\right) + \sum_{i=1}^{n} \alpha_i \left(1 - y^{(i)}\left(\left(\sum_{j=1}^{n} \alpha_j y^{(j)} x^{(j)}\right)^T x^{(i)} + \theta_0\right)\right)$$

$$= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i y^{(i)} \alpha_j y^{(j)} \left(x^{(i)T} x^{(j)}\right) + \sum_{i=1}^{n} \alpha_i \left(1 - y^{(i)}\left[\sum_{j=1}^{n} \alpha_j y^{(j)} x^{(j)T} x^{(i)} + \theta_0\right]\right)$$

Let us assume $\theta_0 = 0$. [General case of $\theta_0 \neq 0$ is skipped]

$\boxed{x^T y = y^T x}$

$$\sup_{\alpha} \underbrace{\mathcal{L}\left(\sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}, \alpha\right)}_{\mathcal{L}(\alpha)} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(i)T} x^{(j)}\right) + \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(j)T} x^{(i)}\right)$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(j)T} x^{(i)}\right)$$

Dual SVM:

$$\sup_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(i)^T} x^{(j)} \right)$$

Predict output on a new value.

$$\theta = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

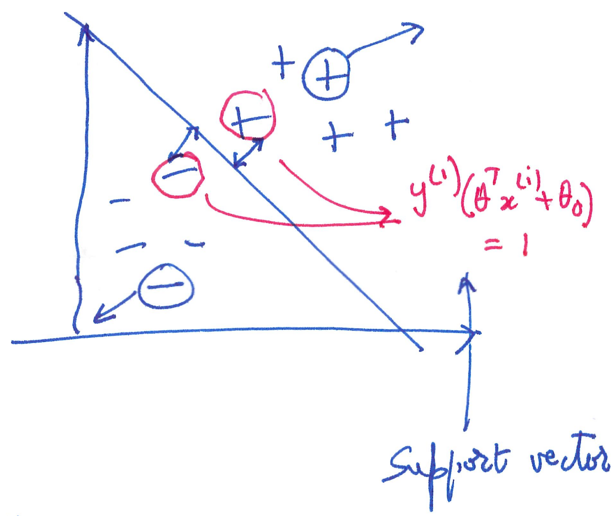
Compute $\theta^T x$.

$$= \left(\sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} \right)^T x.$$

$$= \sum_{i=1}^{n} \alpha_i y^{(i)} \left(x^{(i)^T} x \right).$$

$$\min \|\theta\|^2$$

s.t $\quad y^{(i)} \left(\theta^T x^{(i)} + \theta_0 \right) \geq 1 \quad \forall i$



$$\sup_{\alpha} \left(\frac{1}{2} \|\theta\|^2 + \sum_{i=1}^{n} \alpha_i \underbrace{\left(1 - y^{(i)} \left(\theta^T x^{(i)} + \theta_0 \right)\right)}_{\leq 0} \right)$$

$$\leq \frac{1}{2} \|\theta\|^2$$

If $\alpha_i > 0 \Rightarrow \underline{1 - y^{(i)} \left(\theta^T x^{(i)} + \theta_0 \right)} = 0$

$$\theta = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = \sum_{i:\, y^{(i)}(\theta^T x^{(i)} + \theta_0) = 1} \alpha_i y^{(i)} x^{(i)}$$

$x^T y = \langle x, y \rangle$ [dot product / inner product]

(x, y are vectors)

$$\sup_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \langle x^{(i)}, x^{(j)} \rangle$$

Prediction on x

Compute $\sum_{i=1}^{n} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle$