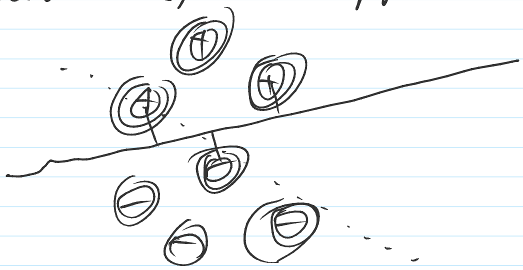
Review 9/26: Support Vector Machines



$$\min_{w,b,\xi\geq 0} \frac{1}{2} w\cdot w + C\sum_{i=1}^{m} \xi_i$$

$$y_1(w\cdot x_1 + b) \geq 1 - \xi_1$$
$$\vdots$$
$$y_m(w\cdot x_m + b) \geq 1 - \xi_m$$

① separte
② large margin $\gamma = \frac{1}{\|w\|}$
③ soften margin: $\sum \xi_i$ is upper bound on # of training errors

---

$$\min_{w,b,\xi\geq 0} \frac{1}{2} w\cdot w + 10 \sum \xi_i$$

s.t. $y_1(w\cdot x_1 + b) \geq 1 - \xi_1$
$y_2(w\cdot x_2 + b) \geq 1 - \xi_2$
$y_3(w\cdot x_3 + b) \geq 1 - \xi_3$
$y_4(w\cdot x_4 + b) \geq 1 - \xi_4$

$$1\underbrace{\left(\begin{pmatrix}1\\0\\0\\0\end{pmatrix}\cdot\begin{pmatrix}1\\0\\0\\0\end{pmatrix}+0\right)}_{1} \geq 1-0$$

$$-1\underbrace{\left(\quad \cdots \quad\right)}_{1} \geq 1-2$$

① $errs(h_{w_1b_1}) = 2$ and $\sum \xi_i = 4$
$\|w_1\| = \sqrt{w_1\cdot w_2} = \sqrt{2} \rightarrow \gamma_1 = \frac{1}{\sqrt{2}} = \frac{1}{\|w_1\|} = 0.71$
→ underfitting
② $errs(h_{w_2b_2}) = 0$ and $\sum \xi_i = 0$
$\|w_2\| = \sqrt{4} \rightarrow \gamma_2 = \frac{1}{2}$
③ $errs(h_{w_3b_3}) = 0$ and $\sum \xi_i = 0$
$\|w_3\| = \sqrt{2} \rightarrow \gamma_3 = \frac{1}{\sqrt{2}} = 0.71$
④ $errs(h_{w_4b_4}) = 0$ and $\sum \xi_i = 0$
$\|w_4\| = \sqrt{3} \rightarrow \gamma_4 = \frac{1}{\sqrt{3}} = 0.58$
⑤ $errs(h_{w_5b_5}) = 0$ and $\sum \xi_i = 0$
$\|w_5\| = \sqrt{1.815} \rightarrow \gamma_5 = 0.74$
⑥ $\|w_6\| = \sqrt{1.33} \rightarrow \gamma_6 = 0.87$

$$w_{final} = 0 + y_1 x_1 + y_5\cdot x_5 + y_{18}\cdot x_{18} + y_5 x_5 + y_{23} x_{23}$$

Updates: 1, 5, 18, 5, 23

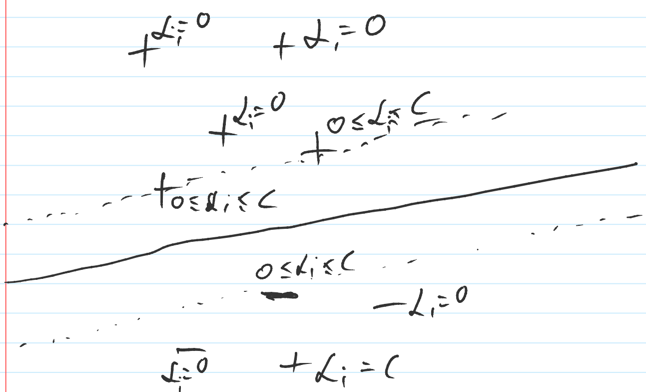
$$\rightarrow w_k = \sum_{j=1}^{m} \text{"Number of updates on } (x_j, y_j)\text{"} \cdot y_j \cdot x_j$$

Idea:

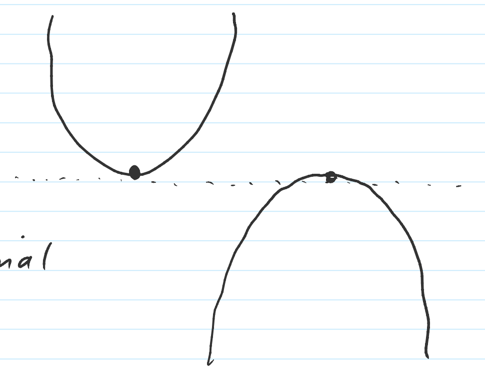
- Do not store w, but record update steps
- Store α_i to contain "Number of updates on (x_i, y_i)"

$$(w_k \cdot x_i) = \underbrace{\left[\sum_{j=1}^{m} \alpha_j y_j x_j\right]}_{w_k} \cdot x_i = \sum_{j=1}^{m} \alpha_j y_j (x_j \cdot x_i)$$

Two equivalent representation
- Primal: w, b
- Dual: $\alpha_1 .. \alpha_m$ and $(x_1, y_1) \dots (x_m, y_m)$



Computing b:
find some (x_i, y_i) with $0 < \alpha_i < C$:
$y_i(w \cdot x_i + b) = 1 \rightarrow b = y_i - x_i \cdot w$



Theorem: If $P(w^*, b^*, \xi^*)$ is solution of Primal and $D(\alpha^*)$ is solution of Dual, then
$P(w^*, b^*, \xi^*) = D(\alpha^*)$

Theorem: The leave-one-out error of an SVM is bounded by

$$err_{loo}(SVM) = \frac{\#\text{ of Support Vectors}}{n}$$

$\rightarrow$ small number of Support Vectors means that estimated generalization error is low!