Lecture 9/24: Optimal Hyperplanes and SVMs

Tuesday, September 24, 2019 2:38 PM

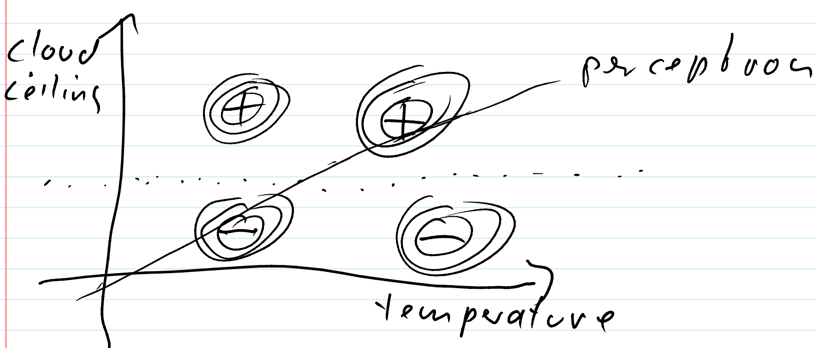
Review of Sep 19

Linear classifier $h_w(x) = \text{sign}(w \cdot x + b)$

Perceptron: Learn a zero training error h_w, if training data is separable

Convergence bound: it take $\frac{R^2}{\gamma^2}$ updates

Optimal hyperplanes



Computing the optimal hyperplane

Training sample $S = ((x_1, y_1) \dots (x_m, y_m))$

Requirement 1: Zero training error!

$$\forall (x_i, y_i) : y_i (w \cdot x_i + b) > 0$$

Requirement 2: Maximum distance to closest training examples.

$$\max_{w,b} \gamma \quad \text{with } \gamma = \min_i \left| \frac{1}{\|w\|} (w \cdot x_i + b) \right|$$

→ Requirement 1 and Requirement 2

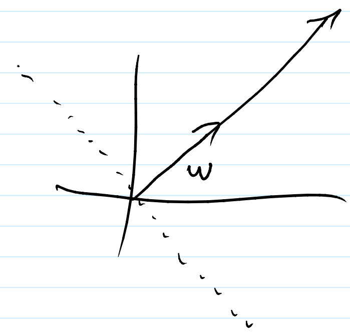
$$\max_{w,b} \gamma \quad \text{with } \gamma = \min_i \left[\frac{y_i}{\|w\|} (w \cdot x_i + b) \right]$$

Simplify:

- Write min as set of constraints

$$\max_{w,b} \gamma$$
$$\text{s.t. } \forall i: \frac{y_i}{\|w\|}(w \cdot x_i + b) \geq \gamma$$

– length of w does not influence solution.
So, we fix $\|w\| = \frac{1}{\gamma}$



$$\max_{w,b} \frac{1}{\|w\|}$$
$$\text{s.t. } \forall i: \frac{y_i}{\cancel{\|w\|}}(w \cdot x_i + b) \geq \frac{1}{\cancel{\|w\|}}$$

– simplify further

$$\min_{w,b} \|w\| = \sqrt{w \cdot w}$$
$$\text{s.t } \forall i: y_i(w \cdot x_i + b) \geq 1$$

→ At solution $\gamma = \frac{1}{\|w\|}$ is geometric margin

Soft-Margin SVM

slack variables at the solution of OP and training errors

$\xi_i \geq 1 \iff y_i(w \cdot x_i + b) \leq 0$ (error)

$0 < \xi_i < 1 \iff 0 < y_i(w \cdot x_i + b) < 1$ (correct, but inside margin)

$\xi_i = 0 \iff y_i(w \cdot x_i + b) \geq 1$ (correct with sufficient margin)