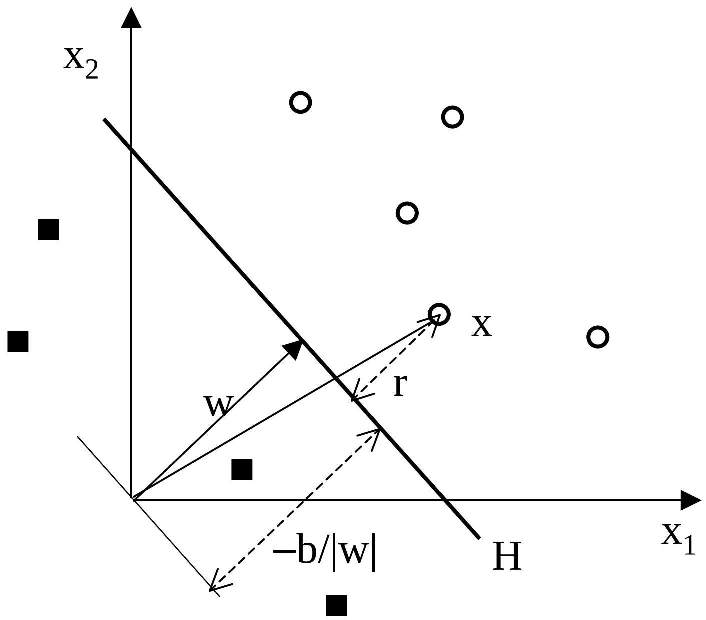
Lecture 17

SVM (I): Basic Formulation

Outline

- Linear pattern classifiers and optimal hyper-plane
- Optimization problem formulation
- Statistical properties of optimal hyper-plane

Linear Hyper-plane Classifier



Given: $\{(x_i, d_i); i = 1 \text{ to } N, d_i \in \{+1, -1\}\}$.
A linear hyper-plane classifier is a hyper-plane consisting of points x such that

$$H = \{x|\ g(x) = w^T x + b = 0\}$$

g(x): a discriminant function!.

For x in the side of o : $w^T x + b \geq 0$; $d = +1$;

For x in the side of ■ : $w^T x + b \leq 0$; $d = -1$.

Distance from x to H: $r = w^T x/|w| - (-b/|w|) = g(x)/|w|$

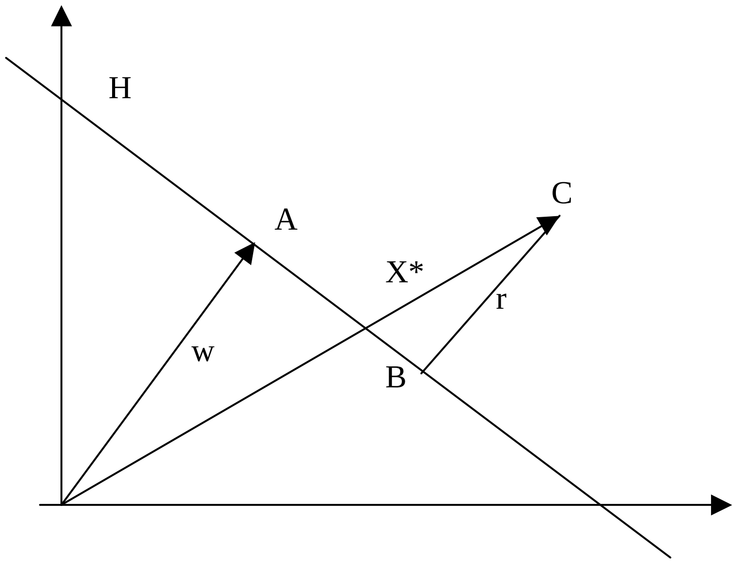
Distance from a Point to a Hyper-plane

The hyper-plane H is characterized by

$$w^T x + b = 0 \qquad (*)$$

w: normal vector perpendicular to H.

(*) says any vector x on H that project to w will have a length of $\overline{OA} = -b/|w|$.



Consider a special point C corresponding to vector x^*. Its magnitude of projection onto vector w is

$$w^T x^*/|w| = \overline{OA} + \overline{BC}.$$

Or equivalently,

$$w^T x^*/|w| = -b/|w| + r$$

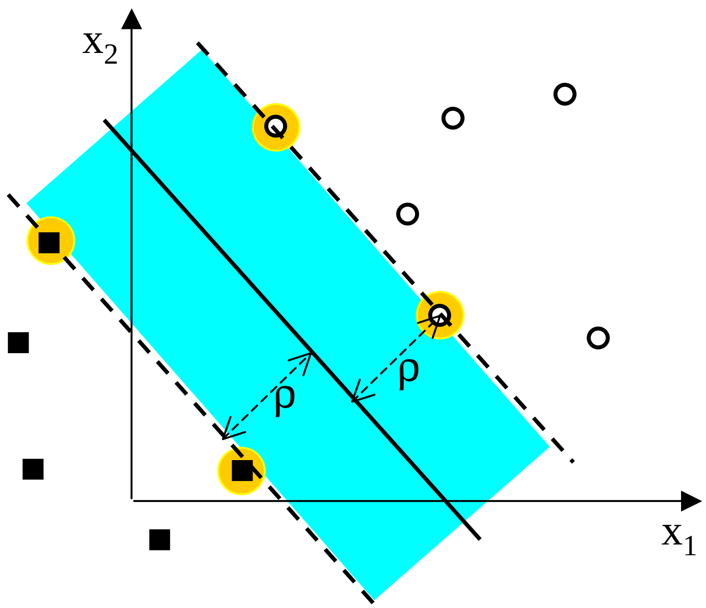
Hence

$$r = (w^T x^* + b)/|w| = g(x^*)/|w|$$

If x* is on the other side of H (same side as the origin), then

$$r = -(w^T x^* + b)/|w| = -g(x^*)/|w|$$

Optimal Hyper-plane: Linearly Separable Case



- Optimal hyper-plane should be in the center of the gap.
- Support Vectors – Samples on the boundaries. Support vectors alone can determine optimal hyper-plane.
- Question: How to find optimal hyper-plane?

For $d_i = +1$, $g(x_i) = w^T x_i + b \geq \rho|w| \Rightarrow w_o^T x_i + b_o \geq 1$

For $d_i = -1$, $g(x_i) = w^T x_i + b \leq -\rho|w| \Rightarrow w_o^T x_i + b_o \leq -1$

Separation Gap

For x_i being a support vector,

For $d_i = +1$, $g(x_i) = w^T x_i + b = \rho|w|$ $\Rightarrow$ $w_o^T x_i + b_o = 1$

For $d_i = -1$, $g(x_i) = w^T x_i + b = -\rho|w|$ $\Rightarrow$ $w_o^T x_i + b_o = -1$

Hence $w_o = w/(\rho|w|)$, $b_o = b/(\rho|w|)$.

But distance from x_i to hyper-plane is $\rho = g(x_i)/|w|$.

Thus $w_o = w/g(x_i)$, and $\rho = 1/|w_o|$.

The maximum distance between the two classes is

$$2\rho = 2/|w_o|.$$

The objective is to find w_o, b_o to minimize $|w_o|$ (so that ρ is maximized) subject to the constraints that

$$w_o^T x_i + b_o \geq 1 \text{ for } d_i = +1; \text{ and } w_o^T x_i + b_o \leq 1 \text{ for } d_i = -1.$$

Combine these constraints, one has:

$$d_i \bullet (w_o^T x_i + b_o) \geq 1$$

Quadratic Optimization Problem Formulation

Given $\{(x_i, d_i); i = 1 \text{ to } N\}$, find w and b such that

$$\phi(w) = w^T w/2$$

is minimized subject to N constraints

$$d_i \bullet (w^T x_i + b) \geq 1; \quad 1 \leq i \leq N.$$

<u>Method of Lagrange Multiplier</u>

$$J(W,b,\alpha) = \phi(W) - \sum_{i=1}^{N} \alpha_i \left[d_i \left(W^T x_i + b \right) - 1 \right]$$

Set $$\frac{\partial J(W,b,\alpha)}{\partial W} = 0 \quad \Rightarrow W = \sum_{i=1}^{N} \alpha_i d_i x_i$$

$$\frac{\partial J(W,b,\alpha)}{\partial b} = 0 \quad \Rightarrow \sum_{i=1}^{N} \alpha_i d_i = 0$$

Optimization (continued)

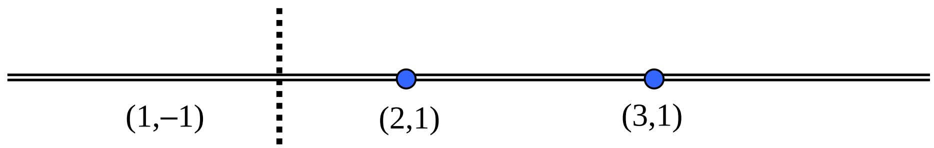
The solution of Lagrange multiplier problem is at a saddle point where the minimum is sought w.r.t. w and b, while the maximum is sought w.r.t. α_i.

<u>(Karush)-Kuhn-Tucker Condition</u>: at the saddle point,

$$\alpha_i[d_i (w^T x_i + b) - 1] = 0 \qquad \text{for } 1 \le i \le N.$$

- If x_i is NOT a support vector, the corresponding α_i = 0!
- Hence, only support vector will affect the result of optimization!

A Numerical Example



3 inequalities:

$$1\bullet w + b \le -1; \quad 2\bullet w + b \ge +1; \quad 3\bullet w + b \ge +1$$

$$J = w^2/2 - \alpha_1(-w-b-1) - \alpha_2(2w+b-1) - \alpha_3(3w+b-1)$$

$$\partial J/\partial w = 0 \Rightarrow w = -\alpha_1 + 2\alpha_2 + 3\alpha_3$$

$$\partial J/\partial b = 0 \Rightarrow 0 = \alpha_1 - \alpha_2 - \alpha_3$$

Kuhn-Tucker condition implies:

(a) $\alpha_1(-w-b-1) = 0$ (b) $\alpha_2(2w+b-1) = 0$ (c); $\alpha_3(3w + b -1) = 0$

Later, we will see the solution is $\alpha_1 = \alpha_2 = 2$ and $\alpha_3 = 0$. This yields $w = 2$, $b = -3$.

Hence the solution of decision boundary is:

$$2x - 3 = 0. \text{ or } x = 1.5!$$

This is shown as the dash line in above figure.

Primal/Dual Problem Formulation

Given a constrained optimization problem with a convex cost function and linear constraints; a *dual problem* with the Lagrange multipliers providing the solution can be formulated.

<u>*Duality Theorem (Bertsekas 1995)*</u>

(a) If the primal problem has an optimal solution, then the dual problem has an optimal solution with the same optimal values.

(b) In order for w_o to be an optimal solution and α_o to be an optimal dual solution, it is necessary and sufficient that w_o is feasible for the primal problem and

$$\Phi(w_o) = J(w_o,b_o,\alpha_o) = \mathrm{Min}_w\, J(w,b_o,\alpha_o)$$

Formulating the Dual Problem

$$J(w,b,\alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i d_i w^T x_i - b \sum_{i=1}^{N} \alpha_i d_i + \sum_{i=1}^{N} \alpha_i$$

At the saddle point, we have $W = \sum_{i=1}^{N} \alpha_i d_i x_i$ and $\sum_{i=1}^{N} \alpha_i d_i = 0$, substituting these relations into above, then we have the

Dual Problem

Maximize $Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j$

Subject to: $\sum_{i=1}^{N} \alpha_i d_i = 0$ and $\alpha_i \ge 0$ for i = 1, 2, …, N.

Note $Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \begin{bmatrix} \alpha_1 d_1 & \cdots & \alpha_N d_N \end{bmatrix} \begin{bmatrix} x_1^2 & \cdots & x_1 x_N \\ \vdots & \ddots & \vdots \\ x_N x_1 & \cdots & x_N^2 \end{bmatrix} \begin{bmatrix} \alpha_1 d_1 \\ \vdots \\ \alpha_N d_N \end{bmatrix}$

Numerical Example (cont'd)

$$Q(\alpha) = \sum_{i=1}^{3} \alpha_i - \frac{1}{2}\begin{bmatrix} -\alpha_1 & \alpha_2 & \alpha_3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} \begin{bmatrix} -\alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

or $Q(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - [0.5\alpha_1^2 + 2\alpha_2^2 + 4.5\alpha_3^2 - 2\alpha_1\alpha_2 - 3\alpha_1\alpha_3 + 6\alpha_2\alpha_3]$

subject to constraints: $-\alpha_1 + \alpha_2 + \alpha_3 = 0$, and

$\alpha_1 \geq 0$, $\alpha_2 \geq 0$, and $\alpha_3 \geq 0$.

Use Matlab™ Optimization tool box command:

x=fmincon('qalpha',X0, A, B, Aeq, Beq)

The solution is $[\alpha_1 \ \ \alpha_2 \ \ \alpha_3] = [2 \ \ 2 \ \ 0]$ as expected.