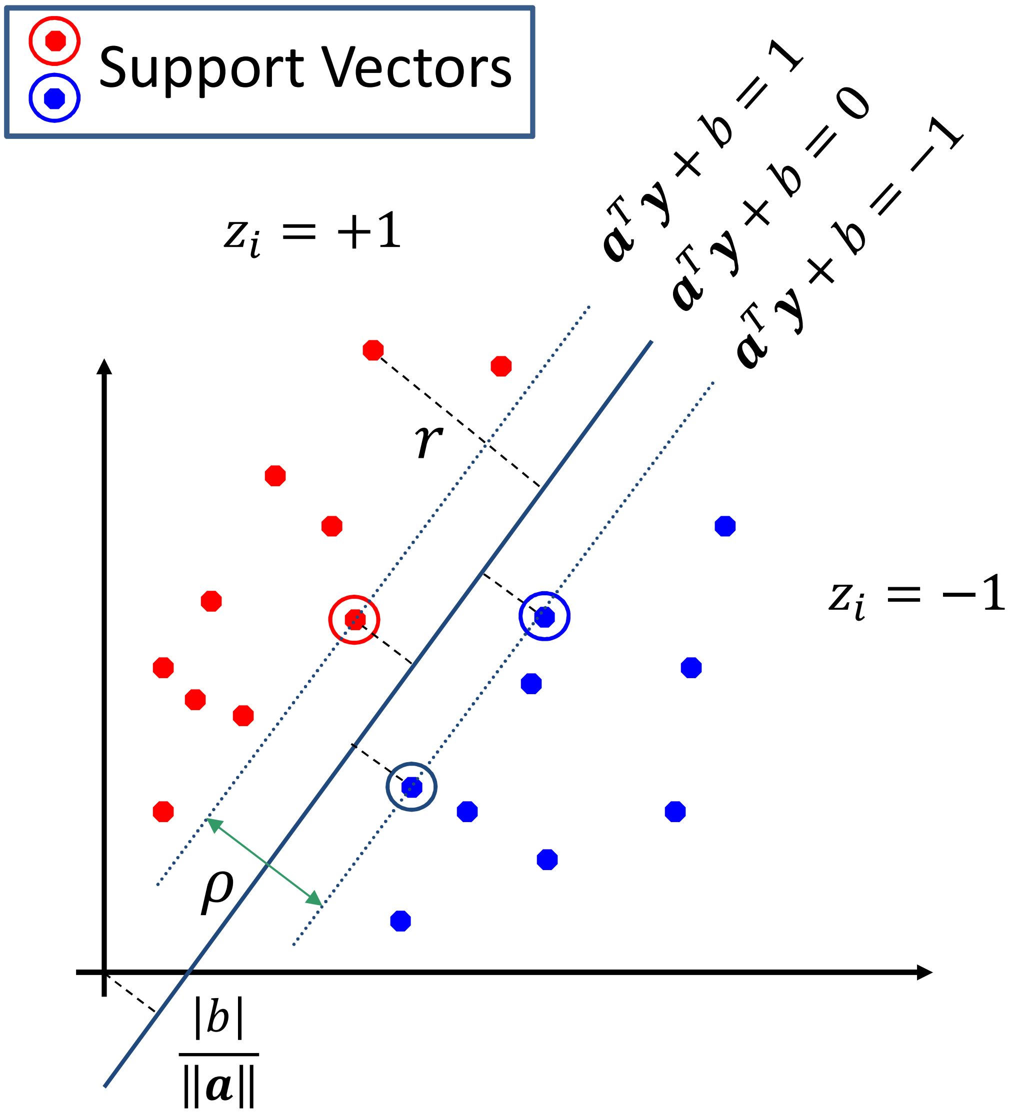
Statistical Methods in Artificial Intelligence
CSE471 - Monsoon 2016 : Lecture 16

Avinash Sharma
CVIT, IIIT Hyderabad

Lecture Plan

- Support Vector Machine (SVM)
  - Primal Formulation
  - Dual Formulation Derivation
  - Soft Margin SVM Derivation
- Transductive, Multi-category and Kernel SVM (Next Class)

Maximum Margin Classification



For points on boundary

$$r_i = \frac{\boldsymbol{a}^T\boldsymbol{y}_i + b}{\|\boldsymbol{a}\|}, z_i = 1$$

$$r_j = \frac{\boldsymbol{a}^T\boldsymbol{y}_j + b}{\|\boldsymbol{a}\|}, z_j = -1$$

$$z_k(\boldsymbol{a}^T\boldsymbol{y}_k + b) = 1$$

$$\rho = r_i - r_j = \frac{2}{\|\boldsymbol{a}\|}$$

Let $|b| \|\boldsymbol{a}\| = constant$

Linear Support Vector Machine

- Primal Formulation:

  Maximize the margin

$$\arg\max_{\boldsymbol{a}} \left(\frac{2}{\|\boldsymbol{a}\|} \right)$$

such that $z_i(\boldsymbol{a}^T \boldsymbol{y}_i + b) \geq 1 \quad \forall i \in \{1, \dots, n\}$

Or,

$$\arg\min_{\boldsymbol{a},b} (\|\boldsymbol{a}\|^2) = \arg\min_{\boldsymbol{a}} \left(\frac{1}{2} \boldsymbol{a}^T \boldsymbol{a} \right)$$

such that $z_i(\boldsymbol{a}^T \boldsymbol{y}_i + b) \geq 1 \quad \forall i \in \{1, \dots, n\}$

Linear Support Vector Machine

- The Primal Formulation optimize a *quadratic* function subject to *linear* constraints.

- Quadratic optimization problems are a well-known class of mathematical programming problems for which several (non-trivial) algorithms exist.

- The solution involves constructing a *dual problem* where a *Lagrange multiplier* α_i is associated with every inequality constraint in the primal (original) problem

Linear Support Vector Machine

- Dual Formulation:

$$\arg \min_{\boldsymbol{a},b} \max_{\alpha_1,\ldots,\alpha_n} \left\{ \frac{1}{2} \boldsymbol{a}^T \boldsymbol{a} - \sum_{i=1}^{n} \alpha_i (z_i (\boldsymbol{a}^T \boldsymbol{y}_i + b) - 1) \right\}$$

such $\alpha_i \geq 0 \quad \forall i \in \{1, \ldots, n\}$

Or,

$$\arg \max_{\alpha_1,\ldots,\alpha_n} \sum_{k=1}^{n} \alpha_k - \frac{1}{2} \sum_{k=1,j=1}^{n} \alpha_k \alpha_j z_k z_j \boldsymbol{y}_k{}^T \boldsymbol{y}_j$$

such that $\sum_{k=1}^{n} \alpha_k z_k = 0$ and $\alpha_k \geq 0 \quad \forall k \in \{1, \ldots, n\}$

Linear Support Vector Machine

- Given a solution $\alpha_1, \dots, \alpha_n$ to the dual problem, solution to the primal is:

$$\boldsymbol{a} = \sum_{j=1}^{n} \alpha_j z_j \boldsymbol{y}_j \text{ and } b_k = z_k - \sum_{j=1}^{n} \alpha_j z_j \boldsymbol{y}_j{}^T \boldsymbol{y}_k \text{ for } \forall \alpha_k > 0$$

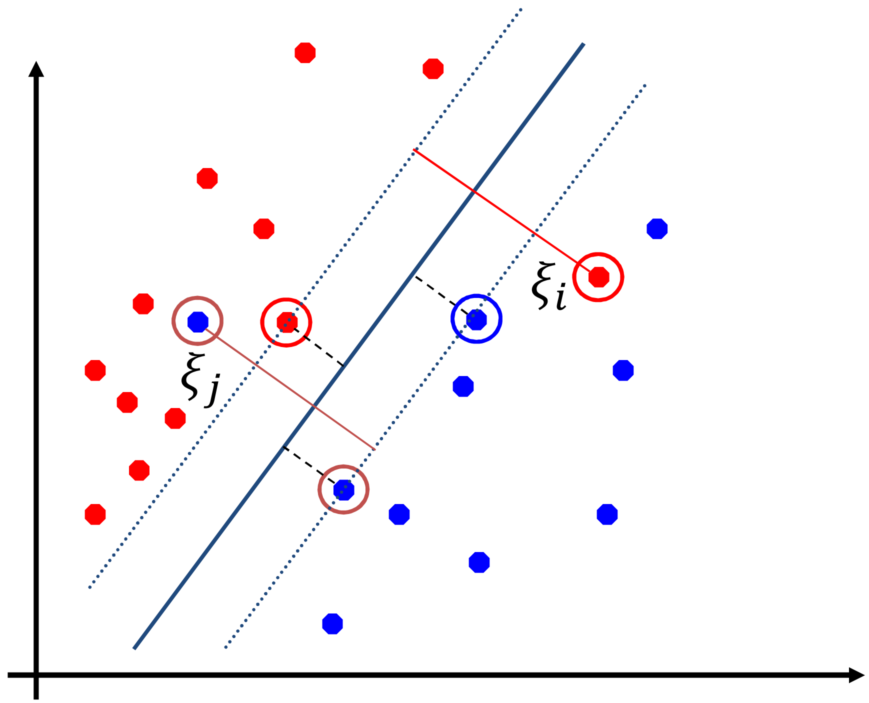
$$b = mean([b_1, \dots, b_k, \dots, b_m])$$

- Each non-zero α_k indicates that corresponding $\boldsymbol{y}_k$ is a support vector.

- The classifying function is:

$$f(\boldsymbol{y}) = \sum_{j=1}^{n} \alpha_j z_j \boxed{\boldsymbol{y}_j{}^T \boldsymbol{y}} + b$$

Soft Margin SVM



Let $\xi_i \geq 0 \;\; \forall i$

$$z_i(\boldsymbol{a}^T \boldsymbol{y}_i + b) \geq 1 - \xi_i$$

Soft Margin SVM

- Primal Formulation:

$$\arg\min_{\boldsymbol{a},\xi,b}(\frac{1}{2}\boldsymbol{a}^T\boldsymbol{a} + C\sum_{i=1}^{n}\xi_i)$$

such that $z_i(\boldsymbol{a}^T\boldsymbol{y}_i + b) \geq 1 - \xi_i$ and $\xi_i \geq 0 \quad \forall i \in \{1, \ldots, n\}$

- Dual Formulation:

$$\arg\max_{\alpha_1,\ldots,\alpha_n}\sum_{k=1}^{n}\alpha_k - \frac{1}{2}\sum_{k=1,j=1}^{n}\alpha_k\alpha_j z_k z_j \boldsymbol{y}_k{}^T\boldsymbol{y}_j$$

such that $\sum_{k=1}^{n}\alpha_k z_k = 0$ and $C \geq \alpha_k \geq 0 \quad \forall k \in \{1, \ldots, n\}$

Soft Margin SVM

- Given a solution $\alpha_1, \ldots, \alpha_n$ to the dual problem, solution to the primal is:

  $\boldsymbol{a} = \sum_{j=1}^{n} \alpha_j z_j \boldsymbol{y}_j$ and $b_k = z_k(1 - \xi_k) - \sum_{j=1}^{n} \alpha_j z_j {\boldsymbol{y}_j}^T \boldsymbol{y}_k$

  for $\forall k \; C \geq \alpha_k \geq 0$

- The classifying function is:

$$f(\boldsymbol{y}) = \sum_{j=1}^{n} \alpha_j z_j \; {\boldsymbol{y}_j}^T \boldsymbol{y} \; + b$$

- Parameter C act as overfitting knob: “trades off” the relative importance of maximizing the margin and fitting the training data.

Reference Material

- www.cs.utexas.edu/~mooney/cs391L/slides/svm.ppt
- http://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf
- www-labs.iro.umontreal.ca/~pift6080/H09/documents/papers/svm_tutorial.ppt
- https://en.wikipedia.org/wiki/Probably_approximately_correct_learning
- https://en.wikipedia.org/wiki/Support_vector_machine