Lecture 22 (More on SVMs)

1 Handling Noise in the Data

1. When the data is not linearly separable (highly practical scenario)
2. Allow for some *slack* for each point, ϵ_i
3. We modify the problem by adding a term $c \cdot \sum_{i=1}^{m} \epsilon_i$ to the primal problem of SVMs
4. Hard SVMs completely separate the classes and soft SVMs allow for flexibility and hence choosing a *better* plane
5. Hard and soft SVMs are same when $c \to \infty$

2 Soft SVMs

$$L(w, b, \alpha, \gamma) = \frac{1}{2} w^T w + c \sum_{i=1}^{m} \epsilon_i + \sum_{i=1}^{m} \alpha_i (1 - y^i (w^T x^i + b) - \epsilon_i) + \sum_{i=1}^{m} \gamma_i (-\epsilon_i)$$

On equating the gradients wrt w, b, ϵ to 0, we get:

$$w = \sum_{i=1}^{m} \alpha_i y^i x^i$$

$$\sum_{i=1}^{m} \alpha_i y^i = 0$$

$$\alpha_i + \gamma_i = c$$

$$\alpha_i \geq 0, \gamma_i \geq 0$$

From complementary slackness, we have three cases:

1. $\alpha_i = 0$ then $\gamma_i = c \implies \epsilon_i = 0, y^i (w^T x^i + b) \geq 1$
   These points don't contribute to the SVM
2. $\gamma_i = 0$ then $\alpha_i = c \implies y^i (w^T x^i + b) = 1 - \epsilon_i$
   There points are inside the margin
3. $0 < \alpha_i < c \wedge 0 < \gamma_i < c$, then $\epsilon_i = 0, y^i (w^T x^i + b) = 1$
   These points are on the margin

The dual problem is given as:

$$\max_{\alpha, 0 \leq \alpha_i \leq c} \left[\sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y^i y^j (x^i)^T x^j \right]$$

The only difference is that α_i has an upper bound. To read about these *box constraints*, read the notes :) (won't be discussed).