Lecture 21 (KKT Conditions)

1 Karush–Kuhn–Tucker Conditions

1. These conditions are necessary and sufficient for the the primal-dual equality
2. Gradient vanishes at optimal parameters
$$\nabla_v L(w, \alpha, \beta)|_{v^*} = 0$$
for $v = w^*, \alpha^*, \beta^*$
3. Primal and dual feasibility exists
4. Complementary slackness
$$\alpha_i g_i(w^*) = 0, \forall i \in \{1, \ldots, m\}$$

2 (new) SVM Objective

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

The SVM dual is given by:
$$\max_{\alpha, \alpha \geq 0} [\underset{w,b}{min} [L(w, b, \alpha)]]$$

Now, we compute the gradient of the Lagrangian:
$$\nabla_w L(w, b, \alpha) = w - \sum_{i=1}^{m} \alpha_i y^i x^i$$

Equating the gradient to 0, we get:
$$w = \sum_{i=1}^{m} \alpha_i y^i x^i$$

Therefore we see that only *active* constraints play a role. Now we compute gradient wrt b and equate it to 0:
$$\nabla_b L(w, b, \alpha) = - \sum_{i=1}^{m} \alpha_i y^i = 0$$

Now substituting these two conditions and eliminating w, b from $L(w, b, \alpha)$, we get:
$$w(\alpha) = \theta_D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y^i y^j x^{iT} x^j$$

Now our dual objective is:

$$\max_{\alpha,\alpha\geq 0,\alpha^T Y=0}[w(\alpha)]$$

1. We can now solve this using Block Coordinate Descent - Sequential Minimal Optimisation [ofc you have to read this by self and it won't be discussed :)]
2. This algo optimises two variables at a time (keeping the other fixed)
3. This simplifies it to a quadratic expression in a single variable by using the second constraint and we are left with the simple constraint of $\alpha_i \geq 0$
4. We now find b as (solved by looking at the inequation of g_i):

$$\frac{-1}{2}[\max_{y^i=-1}(w^T x^i) + \min_{y^i=1}(w^T x^i)]$$

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