Singapore University of Technology and Design
Information Systems Technology and Design
50.007 Machine Learning
Support Vector Machines (Supplement)

Support vector machines (with offset)

As we had discussed before, maximum margin linear separators can be found by solving the following primal quadratic programming problem

$$\text{(primal)} \qquad \min \frac{1}{2}\|\theta\|^2 \ \text{ subject to } \ y^{(t)}(\theta \cdot x^{(t)} + \theta_0) \geq 1, \ \ t = 1, \ldots, n \tag{1}$$

If we wish to solve the problem in the dual, i.e., represent the parameters θ in terms of examples as $\theta = \sum_{t=1}^{n} \alpha_t y^{(t)} x^{(t)}$, we can solve the dual

$$\text{(dual)} \qquad \max \ \sum_{t=1}^{n} \alpha_t - \frac{1}{2} \sum_{t=1}^{n} \sum_{t'=1}^{n} \alpha_t \alpha_{t'} y^{(t)} y^{(t')} (x^{(t)} \cdot x^{(t')}) \tag{2}$$

$$\text{subject to } \ \alpha_t \geq 0, \ \sum_{t=1}^{n} \alpha_t y^{(t)} = 0 \tag{3}$$

where the additional constraint $\sum_{t=1}^{n} \alpha_t y^{(t)} = 0$ pertains to including the offset parameter θ_0 in the primal. But θ_0 does not appear anywhere in the dual. How do we set it? After we solve the dual, we can see that $\hat{\alpha}_t > 0$ for some examples (support vectors) and $\hat{\alpha}_t = 0$ for others (non-support vectors). Since for support vectors, the classification constraints must be satisfied with equality, we have that when $\hat{\alpha} > 0$,

$$y^{(t)}\big(\hat{\theta} \cdot x^{(t)} + \hat{\theta}_0\big) = y^{(t)}\big(\sum_{t'=1}^{n} \hat{\alpha}_{t'} y^{(t')} (x^{(t')} \cdot x^{(t)}) + \hat{\theta}_0\big) = 1 \tag{4}$$

As $y^{(t)} \in \{-1, 1\}$ or $(y^{(t)})^2 = 1$ we can multiply both sides by $y^{(t)}$ and get

$$\hat{\theta}_0 = y^{(t)} - \big(\sum_{t'=1}^{n} \hat{\alpha}_{t'} y^{(t')} (x^{(t')} \cdot x^{(t)})\big) \tag{5}$$

This should hold for all support vectors. However, solving the dual quadratic programming problem numerically does introduce errors and relying on any particular constraint may be unwise. Instead, we can simply take the median of all $\hat{\theta}_0$ estimates from support vector constraints.