Supervision Assignments in MLBI
Lent Term 2018
Set 2

Supervisor : Dionysis Manousakas
dm754@cam.ac.uk

February 21, 2018

1.(*SVMs*)

Assume that the set $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^m \subset \mathbb{R}^2 \times \{-1, 1\}$ of binary examples is strictly linearly separable by a line going trough the origin, that is, there exists $\mathbf{w} \in \mathbb{R}^2$ such that the linear function $f(x) = \mathbf{w}^T\mathbf{x}, \mathbf{x} \in \mathbb{R}^2$ has the property that $y_i f(\mathbf{x}_i) > 0$ for every $i = 1, ..., m$. In this case, a linear separable SVM computes the parameters w by solving the optimisation problem:

$$P1: \quad \min_{\mathbf{w}\in\mathbb{R}^2}\{\frac{1}{2}\mathbf{w}^T\mathbf{w} : y_i\mathbf{w}^T\mathbf{x}_i \geq 1, i = 1, ..., m\}. \tag{1}$$

(a) Show that the vector $\mathbf{w}$ solving problem P1 has the form $\mathbf{w} = \sum_{i=1}^m c_i y_i \mathbf{x}_i$ where $c_1, ..., c_m$ are some nonegative coefficients.

The Laplacian for P1 is the following:

$$L(\mathbf{w}, \mathbf{c}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^m c_i(y_i\mathbf{w}^T\mathbf{x}_i - 1).$$

At the solution we will have

$$0 = \frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^m c_i y_i \mathbf{x}_i,$$

where $c_i \geq 0, i = 1, ..., m$.

(b) Show that the coefficients $c_1, ..., c_m$ in the above formula solve the optimization problem

$$P2: \quad \max\{-\frac{1}{2}\sum_{i,j=1}^m c_i c_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_{i=1}^m c_i : c_j \geq 0, j = 1, ..., m\}. \tag{2}$$

Replacing the coefficient values in the Laplacian we get:

$$
\begin{aligned}
L(\mathbf{w}, \mathbf{c}) =& \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{m} c_i(y_i \mathbf{w}^T \mathbf{x}_i - 1) \\
=& \frac{1}{2}\sum_{i}^{m} c_i y_i \mathbf{x}_i^T \sum_{j}^{m} c_j y_j \mathbf{x}_j - \sum_{i=1}^{m} c_i \big(y_i \sum_{j=1}^{m} c_j y_j \mathbf{x}_j^T \mathbf{x}_i - 1\big) \\
=& \frac{1}{2}\sum_{i,j=1}^{m} c_i c_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i,j=1}^{m} c_i c_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^{m} c_i \\
=& -\frac{1}{2}\sum_{i,j=1}^{m} c_i c_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^{m} c_i
\end{aligned}
$$

At the solution of P1 we maximize the last expression under non-negativity constraints for $c_i, i = 1, ..., m$.

(c) Argue that, if $(\hat{c}_1, ..., \hat{c}_m)$ solves problem P2 and $\hat{\mathbf{w}}$ solves problem P1, then $\hat{\mathbf{w}}^T\hat{\mathbf{w}} = \sum_{i=1}^{m} \hat{c}_i$.

Problem P1 and P2 are dual, with strong duality holding (P1 has convex objective with affine constraint equations). Thus they attain they same solution.