"Advanced Learning Models"
Homework 2

Julien Mairal and Jakob Verbeek

Exercice 1. Dual coordinate ascent algorithms for SVMs

1. We recall the primal formulation of SVMs seen in the class (slide 142).

$$\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i f(\mathbf{x}_i)) + \lambda \|f\|_{\mathcal{H}}^2,$$

and its dual formulation (slide 152)

$$\max_{\boldsymbol{\alpha} \in \mathbb{R}^n} 2\boldsymbol{\alpha}^\top \mathbf{y} - \boldsymbol{\alpha}^\top \mathbf{K} \boldsymbol{\alpha} \quad \text{such that} \quad 0 \leq y_i \alpha_i \leq \frac{1}{2\lambda n}, \quad \text{for all } i.$$

The coordinate ascent method consists of iteratively maximizing the objective with respect to one variable, while fixing the other ones. Assuming that you want to maximize the dual by following this approach. Find (and justify) the update rule for α_j. (note that you need to maximize the objective exactly with respect to one variable, not simply taking a gradient step).

2. Consider now the primal formulation of SVMs with intercept

$$\min_{f \in \mathcal{H}, b \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i (f(\mathbf{x}_i) + b)) + \lambda \|f\|_{\mathcal{H}}^2,$$

Can we still apply the representer theorem? Why? Derive the corresponding dual formulation by using Lagrangian duality. Can we apply the coordinate ascent method to this dual? If yes, what are the update rules?

3. Consider a coordinate ascent method to this dual that consists of updating two variables (α_i, α_j) at a time (while fixing the $n-2$ other variables). What are the update rules for these two variables?

Exercice 2. Kernel examples

Are the following kernels positive definite?

1.

$$\forall x, y \in \mathbb{R} \quad K_1(x,y) = 10^{xy}\,, \quad K_2(x,y) = 10^{x+y}\,.$$

2.

$$\forall x, y \in [0,1) \quad K_3(x,y) = -\log(1-xy)\,.$$

3. Let $\mathcal{X}$ be a set and $f, g : \mathcal{X} \to \mathbb{R}_+$ two non-negative functions:

$$\forall x, y \in \mathcal{X} \quad K_4(x,y) = \min(f(x)g(y), f(y)g(x))$$

Exercice 3. Variational bound on marginal likelihood

Suppose the following mixture distribution $p(x) = \sum_{i=1}^{K} p(z=i)p(x|z=i)$. The entropy of a discrete distribution q is defined as $H(q) = -\sum_{i=1}^{K} q_i \ln q_i$, where we use the shorthand $q_i = q(z=i)$. The Kullback Leibler divergence between distributions p and q is defined as $D(q||p) = \sum_{i=1}^{K} q_i (\ln q_i - \ln p_i)$. Assume all q_i and p_i are strictly positive.

(a) Show that $F \equiv \ln p(x) - D\left(q(z)||p(z|x)\right) \leq \ln p(x)$.

(b) Show that $F = H(q(z)) + \sum_{i=1}^{K} q(z=i) \left[\ln p(z=i) + \ln p(x|z=i)\right]$.

(c) Show that $F = \sum_{i=1}^{K} q(z=i) \left[\ln p(x|z=i)\right] - D\left(q(z)||p(z)\right)$.