Kernel-Based Learning & Multivariate Modeling
MIRI Master - DMKM Master

Lluís A. Belanche
belanche@cs.upc.edu

Problem Set #1, Sept 14, 2016

Problem 1 Ridge regression (primal representation)

In ridge regression, the regularized empirical error to minimize is:

$$E_\lambda(\boldsymbol{w}) = \sum_{n=1}^{N}(t_n - \langle \boldsymbol{w}, \boldsymbol{x}_n \rangle)^2 + \lambda \sum_{i=0}^{d} w_i^2 = \langle \boldsymbol{t} - X\boldsymbol{w}, \boldsymbol{t} - X\boldsymbol{w} \rangle + \lambda \langle \boldsymbol{w}, \boldsymbol{w} \rangle$$

Prove that the solution is given by $\boldsymbol{w}^* = (X^T X + \lambda I_d)^{-1} X^T \boldsymbol{t}$ and derive a form for $f(\boldsymbol{x}) = \langle \boldsymbol{w}^*, \boldsymbol{x} \rangle$.

. . . .

Problem 2 Ridge regression (dual representation)

Prove that the (regularized) solution in **Problem 1** can be also written as:

$$\boldsymbol{w}^* = \sum_{n=1}^{N} \alpha_n \boldsymbol{x}_n$$

In consequence,

$$f(\boldsymbol{x}) = \sum_{n=1}^{N} \alpha_n \langle \boldsymbol{x}_n, \boldsymbol{x} \rangle$$

where $\boldsymbol{\alpha} = (XX^T + \lambda I_N)^{-1} \boldsymbol{t}$.

. . . .

Problem 3 Feature maps and kernels

Given a feature map $\phi : \mathbb{R}^d \to \mathbb{R}^D$, we define its associated kernel function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ as:

$$k(\boldsymbol{u}, \boldsymbol{v}) = \langle \phi(\boldsymbol{u}), \phi(\boldsymbol{v}) \rangle, \qquad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$$

For $\boldsymbol{x} \in \mathbb{R}^d$, consider $\phi(\boldsymbol{x}) = (x_i x_j)_{i,j \in \{1,\ldots,d\}}$.

1. Show that $D = d^2$ and prove that $k(\boldsymbol{u}, \boldsymbol{v}) = \langle \boldsymbol{u}, \boldsymbol{v} \rangle^2$.
2. Calculate the computational cost (as a function of d) of both computing $k(\boldsymbol{u}, \boldsymbol{v})$ and computing $\langle \phi(\boldsymbol{u}), \phi(\boldsymbol{v}) \rangle$ directly.
3. Generalize the previous results to $\phi(\boldsymbol{x}) = (x_{i_1} x_{i_2} \cdots x_{i_q})_{i_1, i_2, \ldots, i_q \in \{1, \ldots, d\}}$.

. . . .

Problem 4 Kernel ridge regression

If we take the simple choice $\phi(\boldsymbol{x}) = \boldsymbol{x}$, $d = D$ and $k(\boldsymbol{u}, \boldsymbol{v}) = \langle \boldsymbol{u}, \boldsymbol{v} \rangle$ (equal to $\boldsymbol{u}^T \boldsymbol{v}$ in this case), apply the Representer Theorem to show that the regularized solution can be written as

$$f(\boldsymbol{x}) = \sum_{n=1}^{N} \alpha_n \langle \boldsymbol{x}_n, \boldsymbol{x} \rangle = \sum_{n=1}^{N} \alpha_n k(\boldsymbol{x}_n, \boldsymbol{x})$$

where the vector of parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^T$ is given by $\boldsymbol{\alpha} = (K + \lambda I_N)^{-1} \boldsymbol{t}$, being $K = (k_{ij})$, with $k_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

. . . .

Problem 5 Representer Theorem and ridge regression

Give a proof for the Representer Theorem as it is sketched in the first set of class slides (we'll see it in a more general setting later on). Apply the Theorem to the case $L(a, b) = (a - b)^2$ (square loss) and ridge regression; you should get the result of Problem 4.

. . . .

Problem 6 Practice play with kernel ridge regression

Consider the function

$$f(x) = 0.5 \frac{\sin(x - a)}{x - a} + 0.8 \frac{\sin(x - b)}{x - b} + 0.3 \frac{\sin(x - c)}{x - c}$$

where $a = 10, b = 50, c = 80$. Generate a dataset of $N = 1052$ examples where the x are equally-spaced in $[0.1, 100]$ and the targets for regression are obtained as $t = f(x) + N(0, 0.05^2)$. Fit standard polynomial regression with some degrees of your choice; then fit kernel ridge regression with the RBF kernel and some σ and λ of your choice, until you are satisfied with the fit. Write a small report (max. 4 pages) with your results.

. . . .