Tutorial 5

Monday 29$^{\text{th}}$ August, 2016

Problem 1. Relation between Penalized Ridge Regression (λ) and Constrained Ridge Regression (θ):

Show that the solution to the *Constrained Ridge Regression* problem

$$\mathbf{w}_{Con} = \underset{\mathbf{w}}{\operatorname{argmin}} \; ||\phi\mathbf{w} - \mathbf{y}||_2^2$$

such that $||\mathbf{w}||_2^2 \leq \xi$

is the same as that to the solution to *Penalized Ridge Regression*.

$$\mathbf{w}_{Pen} = \underset{\mathbf{w}}{\operatorname{argmin}} \; ||\phi\mathbf{w} - \mathbf{y}||_2^2 + \lambda||\mathbf{w}||_2^2$$

for some λ that is a function of ξ.

Hint: You can make convexity assumptions and use KKT conditions.

Problem 2. Consider a data set in which each data point y_i is associated with a weighting factor r_i, so that the sum-square error function becomes

$$\frac{1}{2}\sum_{i=1}^{m} r_i(y_i - w^T\phi(x_i))^2$$

Find an expression for the solution w^* that minimizes this error function. The weights r_i's are known before hand. (Exercise 3.3 of Pattern Recognition and Machine Learning, Christopher Bishop)

Problem 3. In problem 2, we discussed weighted regression. In this problem, we will deal with weighted regression, with the weights obtained using some kernel $K(.,.)$. Given a training set of points $\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_i, y_i), \ldots, (\mathbf{x}_n, y_n)\}$, we predict a regression function $f(x') = (\mathbf{w}^\top\phi(x') + b)$ for each test (or query point) x' as follows:

$$(\mathbf{w}', b') = \underset{\mathbf{w},b}{\operatorname{argmin}} \sum_{i=1}^{n} K(x', x_i)\left(y_i - (\mathbf{w}^\top\phi(x_i) + b)\right)^2$$

1. If there is a closed form expression for $(\mathbf{w}', b')$ and therefore for $f(x')$ in terms of the known quantities, derive it.

2. How does this model compare with linear regression and $k-$nearest neighbor regression? What are the relative advantages and disadvantages of this model?

3. In the one dimensional case (that is when $\phi(x) \in \Re$), graphically try and interpret what this regression model would look like, say when $K(.,.)$ is the linear kernel[1].

Problem 4. Put together the entire story of Support Vector Regression (SVR) one place. You can structure your story along the following lines

1. The motivation behind the basic formualation(s) of SVR with justification for each component of the objective and constraints.

2. The Langrangian function and KKT conditions.

3. Solutions to the KKT conditions and the concept of Support Vectors.

4. The Dual Optimization problem for SVR and the Kernel function

5. Valid kernel functions.

Problem 5. Optional, but could help in ML Project:

For each of the following functions, determine whether a local minimum (maximum) will correspond to a global minimum (maximum). You do not have to prove anything rigorously. You just need to understand the intuitive reason and can consult any web or textual resources.

1. $f(x) = e^x - 1$ on $\mathbb{R}$.

2. $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$.

3. $f(x_1, x_2) = 1/(x_1 x_2)$ on $\mathbb{R}^2_{++}$.

4. $f(x_1, x_2) = x_1/x_2$ on $\mathbb{R}^2_{++}$.

5. $f(x_1, x_2) = x_1^2/x_2$ on $\mathbb{R} \times \mathbb{R}_{++}$.

6. $f(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha})$, where $0 \leq \alpha \leq 1$, on $\mathbb{R}^2_{++}$.

7. Suppose $p \geq 1$ and
$$f(x) = \left(\sum_{i=1}^{n} x_i^P \right)^{1/p}$$
with domain dom $f = \mathbb{R}^n_{++}$ What if $p < 1, p \neq 0$?

[1] Hint: What would the regression function look like at each training data point?