MA2823: Introduction to Machine Learning **Homework 6**

Due November 17, 2017.

Turn in your homework

— as a PDF file (**This means no .docx**)

— named `HW06_<LastName>_<FirstName>.pdf` (no accents)

— at http://tinyurl.com/ma2823-2017-hw

Question

Assume we are given data $\{(\boldsymbol{x}^1, y^1), \dots, (\boldsymbol{x}^n, y^n)\}$ where $\boldsymbol{x}^i \in \mathbb{R}^p$ and $y^i \in \mathbb{R}$, and a parameter $\lambda > 0$. We denote by X the $n \times p$ matrix of row vectors $\boldsymbol{x}^1, \dots, \boldsymbol{x}^n$ and $\boldsymbol{y} = (y^1, \dots, y^n)$.

We are also given a *graph structure* on the features, where vertices are features and edges connect related features. We denote by $\mathcal{E}$ the set of edges of this graph.

The graph-Laplacian-regularized linear regression estimator is defined as:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} ||\boldsymbol{y} - X\boldsymbol{\beta}||_2^2 + \lambda \sum_{(u,v) \in \mathcal{E}} (\beta_u - \beta_v)^2.$$

What does the regularizer $\sum_{(u,v) \in \mathcal{E}} (\beta_u - \beta_v)^2$ enforce (in other words, what is going to be special about the solution)?

Solution: There exists a unique $t \in \mathbb{R}^+$ such that the formulation is equivalent to

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} ||\boldsymbol{y} - X\boldsymbol{\beta}||_2^2 \text{ subject to } \sum_{(u,v) \in \mathcal{E}} (\beta_u - \beta_v)^2 \leq t.$$

If we want to minimize this second term, then for any pair (u, v) of connected features, the quantity $(\beta_u - \beta_v)^2$ should be small. Hence the regularizer forces connected features to have similar weights.