STK-IN4300 Project 1 Exercise 2

steinnhauser

October 2019

We have the expression:

$$F = \sum_{i=1}^{N} g'(w_{old}x_i)^2 \left(\frac{y_i - g(w_{old}x_i)}{g'(w_{old}x_i)} + w_{old}^T x_i - w^T x_i \right)^2 \tag{1}$$

or, gathering the terms in new variables yields

$$F = \sum_{i=1}^{N} \beta_i \left(\alpha_i - x_i^T w_i \right)^2, \tag{2}$$

where

$$\alpha_i = \frac{y_i - g(w_{old}x_i)}{g'(w_{old}x_i)} + x_i^T w_{old}, \tag{3}$$

and

$$\beta_i = g'(w_{old}x_i)^2. \tag{4}$$

The problem is reduced to a *weighted least squares* problem, where β_i represents the weight, and α_i represents the target/objective for component $i \in [1, \ldots, N]$.

This problem can be rewritten as a linear algebra problem; the squared term can be rewritten using the following relation which we know from *ordinary least squares*:

$$\sum_i (y - X\beta)^2 = (y - X\beta)^T (y - X\beta) \tag{5}$$

Rewriting equation 2 involves utilizing this relation in addition to introducing a new way to express the weight term β_i. The way this is done is to set up a diagonal matrix B, where

$$B = \begin{bmatrix} \beta_1 & 0 & 0 & \dots & 0 \\ 0 & \beta_2 & 0 & \dots & 0 \\ 0 & 0 & \beta_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \beta_n \end{bmatrix}. \tag{6}$$

This can now be used to rewrite equation 2 in terms of a linear algebra equation:

$$\sum_{i=1}^{N} \beta_i \left(\alpha_i - x_i^T w_i \right)^2 = \left(\hat{\alpha} - X^T \hat{w} \right)^T B \left(\hat{\alpha} - X^T \hat{w} \right), \tag{7}$$

where the parameters $\hat{\alpha}$ and $\hat{\beta}$ are now vectors comprised of the scalars α_i and β_i, respectively, given by;

$$\hat{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n], \tag{8}$$

$$\hat{\beta} = [\beta_1, \beta_2, \ldots, \beta_n], \tag{9}$$

and X is now a matrix comprised of the vectors x_i, given by:

$$X = [x_1, x_2, \ldots, x_N] \tag{10}$$

Expanding equation 7 yields:

$$F = \left(\hat{\alpha} - X^T\hat{w}\right)^T B \left(\hat{\alpha} - X^T\hat{w}\right) \tag{11}$$

$$= \hat{\alpha}^T B\hat{\alpha} - \hat{\alpha}^T BX^T\hat{w} - \hat{w}^T XB\hat{\alpha} + \hat{w}^T XBX^T\hat{w} \tag{12}$$

by linear algebra multiplication. This is the function which we wish to minimize with respect to the parameter w. To derive the expressions for w which minimize F, the following equation is used:

$$\frac{\partial F}{\partial \hat{w}} = 0 \tag{13}$$

Here, it is assumed that there is only one such extremal solution, and that the solution is a minimum of the function. Inserting the expression for F and taking the derivative with respect to w yields:

$$\frac{\partial}{\partial \hat{w}} \left(\hat{\alpha}^T B\hat{\alpha} - \hat{\alpha}^T BX^T\hat{w} - \hat{w}^T XB\hat{\alpha} + \hat{w}^T XBX^T\hat{w}\right) = 0. \tag{14}$$

This simplifies away the factor which does not depend on w:

$$\frac{\partial}{\partial \hat{w}} \left(-\hat{\alpha}^T BX^T\hat{w} - \hat{w}^T XB\hat{\alpha} + \hat{w}^T XBX^T\hat{w}\right) = 0. \tag{15}$$

A quick dimensional analysis reveals that the terms are scalars, such that:

$$\hat{\alpha}^T BX^T\hat{w} = \left(\hat{\alpha}^T BX^T\hat{w}\right)^T = \hat{w}^T XB\alpha, \tag{16}$$

where $B^T = B$ is a diagonal matrix. This allows us to write equation 15 as:

$$\frac{\partial}{\partial \hat{w}} \left(-2\hat{\alpha}^T BX^T\hat{w} + \hat{w}^T XBX^T\hat{w}\right) = 0. \tag{17}$$

Differentiating the trivial term:

$$-2\hat{\alpha}^T BX^T + \frac{\partial}{\partial \hat{w}} \left(\hat{w}^T XBX^T\hat{w}\right) = 0. \tag{18}$$

The final derivation requires the following matrix calculus identity:

$$\frac{\partial x^T Ax}{\partial x} = 2x^T A \tag{19}$$

This applies to a matrix A which is not a function of $\hat{x}$ and is *symmetric*. In our case, $A = XBX^T$ is a symmetric matrix due to $(XBX^T)^T = XBX^T$, such that the final expression for the minimum of the F function with respect to $\hat{w}$ is then given by:

$$-2\hat{\alpha}^T BX^T + \frac{\partial}{\partial \hat{w}} \left(\hat{w}^T XBX^T\hat{w}\right) = 0 \tag{20}$$

$$-2\hat{\alpha}^T BX^T + 2\hat{w}^T XBX^T = 0, \tag{21}$$

solving for $\hat{w}$:

$$\hat{w}^T XBX^T = \hat{\alpha}^T BX^T \tag{22}$$

$$\hat{w}XBX^T = XB\hat{\alpha} \tag{23}$$

$$\underline{\underline{\hat{w} = \left(XBX^T\right)^{-1} XB\hat{\alpha}}} \tag{24}$$