CS 260 : Machine Learning Algorithm
Homework #4

Ronak Sumbaly
UID - 604591897

October 29, 2015

Problem 1 - Linear Regression with Heterogenous Noise

Log-Likelihood Function

Canonical Setup

Given Linear Regression of the form,

$$y_n = x_b^T \beta + \varepsilon_n$$

Where, $\varepsilon \sim \mathcal{N}(0, \sigma_n)$ and σ_b are not the same all n.

Calculation

$$P(y_n|\beta, x_n) = \frac{exp\left(-\frac{\left(y_n - x_n^T\beta\right)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma_n^2}} \tag{1}$$

Likelihood Function

$$P(D) = \prod_{n=1}^{N} P\ (y_n \mid \beta, x_n) \tag{2}$$

Applying log to the above equation (2) we get,

Log Likelihood Function

$$logP(D) = \sum_{n=1}^{N} log\ p\ (y_n \mid \beta, x_n) \tag{3}$$

$$= \sum_{n=1}^{N} \left[\frac{-\ log\ (2\pi\sigma_n^2)}{2} - \frac{(y_n - x_n^T\ \beta)^2}{2\sigma^2}\right] \tag{4}$$

Equation 4. denotes the log-likelihood function of the data.

Maximum Log-Likelihood Function of β

Differentiating Log-Likelihood function (equation 4.) w.r.t β we get,

$$\frac{\partial \ logP(D)}{\partial \beta} = \sum_{n=1}^{N} \frac{1}{\sigma^2} \ (y_n - x_n^T \ \beta) \ x_n^T \tag{5}$$

To calculate maximum log-likelihood function of β equate equation 5. to 0. Equating we get,

$$\sum_{n=1}^{N} \frac{1}{\sigma_n^2} \ y_n \ x_n^T = \sum_{n=1}^{N} \frac{1}{\sigma_n^2} \ x_n^T \ \beta \ x_n \tag{6}$$

Rearranging to find value for β we get,

$$\boxed{\widehat{\beta}^{\ T} = \left[\ \sum_{n=1}^{N} \ x_n \ x_n^T \ \right]^{-1} \left[\ \sum_{n=1}^{N} \ x_n^T \ y_n \ \right]} \tag{7}$$

Equation 7. denotes the log-likelihood function of the data.

Problem 2 - Linear Regression with Smooth Coefficients

Optimization Problem for β

Canonical Setup

We can encode the natural ordering information for the linear model by introducing a condition that requires the difference $(\beta_i - \beta_{i+1})^2$ cannot be large.

Calculation

To formulate the condition as a regularizer, we define the matrix M $\in \mathbf{R}^n$ as follows,

$$\begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 0 & 1 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

The regularizer can be represented as,

$$Regularizer \ term = ||M \ \beta||_2^2$$

New optimization problem for finding β by combining both this regularization and $L2$ regularization.

$$\boxed{L \ (\beta) = || \ y - X\beta \ ||_2^2 \ + \lambda_1 \ || \ \beta \ ||_2^2 \ + \lambda_2 \ || \ M\beta \ ||_2^2} \tag{8}$$

Equation 8. represents the new optimization problem for finding β.

Optimal β value

Expanding equation 8. as per the rules of L2 regularization we get,

$$L\ (\beta) = ((y^T \cdot y) - (2 \cdot y^T \cdot X \cdot \beta) + (\beta^T \cdot X^T \cdot X \cdot \beta)) + (\lambda_1 \cdot \beta^T \cdot \beta) + (\lambda_2 \cdot \beta^T \cdot M^T \cdot M \cdot \beta) \tag{9}$$

Differentiating equation (9) w.r.t β

$$\frac{\partial L\ (\beta)}{\partial \beta} = -(2 \cdot X^T \cdot y) + (2 \cdot X^T \cdot X \cdot \beta) + (2 \cdot \lambda_1 \cdot \beta) + (2 \cdot \lambda_2 \cdot M^T \cdot M \cdot \beta) \tag{10}$$

Equating equation (10) to find optimal value of β, we get by rearranging,

$$\boxed{\widehat{\beta}\ = ((X^T \cdot X) + (\lambda_1 \cdot I) + (\lambda_2 \cdot M^T \cdot M))^{-1}\ (X^T y)} \tag{11}$$

Equation 11 represents the optimal value of β.

Problem 3 - Linearly Constrained Linear Regression

Canonical Setup

To find the maximum likelihood estimation of β for the specified linear model under the constraint that $A\beta = b$

Lagrangian Multiplier

Consider the problem as an L2 minimization problem we proceed by using the Lagrange's Approach.

$$\mathcal{L}(\beta, \lambda) = ||\ y - X\beta\ ||_2^2\ + \lambda(A\beta - b) \tag{12}$$

Differentiating equation (12) w.r.t β,

$$\frac{\partial L\ (\beta)}{\partial \beta} = ((2 \cdot X^T \cdot X \cdot \beta) - (2 \cdot X^T \cdot y) + (A^T \cdot \lambda)) \tag{13}$$

Rearranging the above equation to get β,

$$\beta = (X^T \cdot X)^{-1} X^T y - \frac{1}{2}(X^T \cdot X)^{-1} A^T \lambda \tag{14}$$

Applying the constraint on the equation (14) by multiplying A on both sides.

$$A\beta = b = A\ (X^T \cdot X)^{-1} X^T y - \frac{1}{2} A\ (X^T \cdot X)^{-1} A^T \lambda \tag{15}$$

Rearranging above equation to get λ,

$$A\ (X^T \cdot X)^{-1} X^T y - b = \frac{1}{2} A\ (X^T \cdot X)^{-1} A^T \lambda \tag{16}$$

$$\lambda = 2(A\ (X^T \cdot X)^{-1} A^T)^{-1} (A\ (X^T \cdot X)^{-1} X^T y - b) \tag{17}$$

Substituting value of λ in equation (14). We get,

$$\boxed{\beta = \left[(X^T \cdot X)^{-1} X^T y\right] - \left[(X^T \cdot X)^{-1} A^T\ (A\ (X^T \cdot X)^{-1} A^T)^{-1} (A\ (X^T \cdot X)^{-1} X^T y - b)\right]} \tag{18}$$

Equation 18 represents the maximum likelihood estimation of β under the constraint.

Problem 4 - Online Learning

Considering the following equations for representation of w_{i+1}

$$w_{i+1} = argmin\ \frac{1}{2} || \ w_{i+1} - w_i \ ||_2^2 \tag{19}$$

Under the constraint of $w^T x_{i+1} y_{i+1} = 0$

By using Lagrange Multiplier and expanding the L2 norm form we get,

$$\mathcal{L}(w, \lambda) = \frac{1}{2} (\ w_{i+1} - w_i \ )^T \ (\ w_{i+1} - w_i \ ) + \lambda y_i x_i^T w \tag{20}$$

Differentiating equation 20 w.r.t **w**, we get,

$$\frac{\partial L \ (w, \lambda)}{\partial w} = (w_{i+1} - w_i) - \lambda x_i y_i \tag{21}$$

Equating above equation (21) to 0 to get **w**

$$w_{i+1} = \lambda x_i y_i + w_i \tag{22}$$

Differentiating equation 21 w.r.t λ

$$\begin{aligned} y_i w_i^T x_i - \lambda y_i y_i x_i^T x_i &= 0 \\ \lambda \ || \ x_i \ ||^2 = y_i w_i^T x_i \qquad \text{Since } y_i^2 &= 1 \\ \lambda &= \frac{y_i w_i^T x_i}{|| \ x_i \ ||^2} \end{aligned}$$

Substituting the value of λ in the above equation we get the new update rule as :

$$\boxed{w_{i+1} = w_i + \frac{w_i^T x_i}{||x_i||_2^2} x_i \tag{23}}$$