STAD80: Homework #2

Harpreet Sumal

Due: Sunday, February, 23rd

Contents

Question 1

Answers:

1. Suppose $\hat{\beta}_a$ is a minimizer for 0.1. Let $C = ||\hat{\beta}_a||_1$ (i.e., the 1-norm for $\hat{\beta}_a$).

Assume that the β minimizer for (0.2) is unique, and let this $\beta = \hat{\beta}_b$.

We also know that the only way $\hat{\beta}_a \neq \hat{\beta}_b$ is if and only if $\hat{\beta} > C$.

Since $C = ||\hat{\beta}_a||_1$ we know that it is possible for $\hat{\beta}_a = \hat{\beta}_b$ by definition.

However, we will assume that $\hat{\beta}_b$ is unique for (0.2)

$\implies ||\hat{\beta}_b||_1 \leq C \leq ||\hat{\beta}_a||_1$

$= \frac{1}{2n}||Y - X\hat{\beta}_b||_2^2 \leq \frac{1}{2n}||\hat{\beta}_a||_2^2$

$\implies \frac{1}{2n}||Y - X\hat{\beta}_b||_2^2 \leq \frac{1}{2n}||Y - X\hat{\beta}_a||_2^2 + \lambda||\hat{\beta}_a||_1$

From lecture we know both parts of the lasso regression are convex,

$\implies$ for all $\theta \in (0, 1)$

$\implies ||Y - X[(1-\theta)\hat{\beta}_a + \theta\hat{\beta}_b]||_2^2 + \lambda||(1-\theta)\hat{\beta}_a + \theta\hat{\beta}_b||_1$

$< ||Y - X\hat{\beta}_a||_2^2 + \lambda_1||\hat{\beta}_a||_1$

Contradiction, as wanted

■

2. First let us re-define some of the variables in (0.3)

Let $\lambda_2||\beta||_2^2 = \lambda[\alpha||\beta||_2^2]$ where $\lambda_2 = \lambda(\alpha)$

Let $\lambda_1||\beta||_1 = (1-\alpha)||\beta||_1$ where $\lambda_1 = (1-\alpha)$

Now by removing the association with both an α and λ let us construct $\tilde{X}, \tilde{y}, \tilde{\lambda}$

$\tilde{X} = \begin{pmatrix} X \\ \sqrt{\lambda_2} \cdot I \end{pmatrix}$. We construct $\tilde{X}$ this way so that when $\tilde{X}\beta$ is calculated, we this creates a new row entry for the value of $\lambda_2||\beta||_2^2$.

$\tilde{y} = \begin{pmatrix} y \\ 0 \end{pmatrix}$ so that our value of $\tilde{X}\beta$ isn't changed.

$\tilde{\lambda} = \lambda_1$

■

Question 2

Answers:

A. We know $log(\lambda(X)) = \beta^T X$, what we want is $log\ L(\beta|(Y|X))$.

$$
\begin{aligned}
L(B|(Y|X)) &= \Pi_{i=1}^{n} \frac{e^{-\lambda(x_i)}[\lambda(x_i)]^{y_i}}{y_i!} \\
&= \Pi_{i=1}^{n} \frac{e^{-e(\beta^T X)}[e^{\beta^T X}]^{y_i}}{y_i!}
\end{aligned}
$$

$log\ L(B|(Y|X)) = -\sum e^{\beta^T X} + \sum y_i \beta^T X - \sum log(y_i)$

■

B. I have no concrete idea of what I'm supposed to do with this quesiton, even just understanding what exactly it is asking me. From my understanding it's that the $E_{\hat{\beta}}$ at the MLE of $\hat{\beta}$ of the given conditional distribution, is equal to $\sum y_i x_i$. However, doing that calculation is something I can't really wrap my head around, albeit I know it should not be too difficult. Hopefully my head's in the right (although very generic) direction.

Question 3

Answers:

1. Let x_1 be a point on the first hyperplane

$\beta^T X + b = 1 \implies b = +1$ since $\beta^T X = 0$. Let this $b = b_1$

$\beta^T X + b = -1 \implies b = -1$ since $\beta^T X = 0$. Let this $b = b_2$

Let L be the line that passes through x_1 in the direction of β

An equation for L is $x_1 + \beta t$ for all $t \in R$

Now let's find an intersection of L on the second plane:

$\beta^T(x_1 + \beta t) = b_2 \Leftrightarrow t = \frac{(b_2 - \beta^T x_1)}{\beta^T \beta} = \frac{(b_2 - b_1)}{\beta^T \beta}$

Therefore the instersection point is $x_2 = x_1 + \beta \frac{(b_2 - b_1)}{\beta^T \beta}$ and the corresponding distance is,

$$
\begin{aligned}
||x_1 - x_2|| &= \frac{|b_2 - b_1|}{\beta^T \beta} ||\beta|| \\
&= \frac{|b_2 - b_1|}{||\beta||_2} \\
&= \frac{|-1-1|}{||\beta||_2} \\
&= \frac{2}{||\beta||_2}
\end{aligned}
$$

■

2. (a) Since the goal of SVM is maximizing the distance between hyperplanes, and the distance between two hyperplanes is represented by a factor of β (as $\frac{2}{||\beta||_2}$). Minimizing β in this case will result in the maximum distance.

Case: $y_i = 1$

$1(\beta^T x_i - b) \geq 1$

$1(0 - b) \geq 1$

$-b \geq 1$

Case: $y_i = -1$

$-1(\beta^T x_i - b) \geq 1$

$-1(0 - b) \geq 1$

$b \geq 1$

These y_i values gives us the corresponding b-values to make our parallel hyperplanes to conduct SVM. Therefore, by constraining in this manor, we are able to find our parallel hyperplanes/b-values.

(b) Any point with $y_i = 0$ will result in an infeasible solution since we will get something of the form:

$0(\beta^T x_i - b) \geq 1$

$0(0 - b) \geq 1$

$0 \geq 1$, which can never be true.

Hence, choose {(0, 0), (0, 1), (0, -1)}

3. If (0.5) holds, which is an example of true SVM, our soft margin, reformulation of SVm will automatically hold by definition by simply redefining our $\hat{\zeta}_i$'s