Penalized Logistic Regression

-Applied Analytics-

APM Chapter 12.15

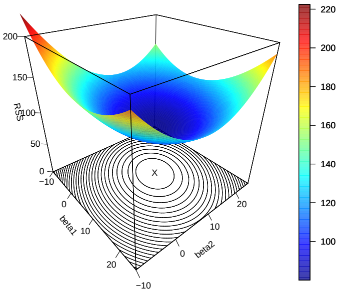
Lecturer: Darren Homrighausen, PhD

Multiple linear regression

Reminder: Multiple linear regression is estimated by finding the minimum of

- the squared error loss
- evaluated on the training observations

  (This is known as the apparent/training error)



Logistic regression

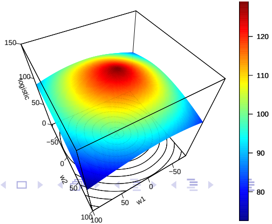
Reminder: Logistic regression is estimated by modeling the probabilities as

$$p = \frac{\exp\{\beta_0 + \sum_{j=1}^{p} \beta_j x_j\}}{1 + \exp\{\beta_0 + \sum_{j=1}^{p} \beta_j x_j\}}.$$

which can also be written as

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$$

The β_0 and coefficient vector β gets estimated by maximizing the 'likelihood' given by this model



Logistic regression

Logistic regression maximizes the (log) likelihood:

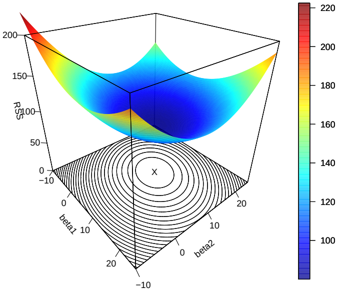
(Using $Y_i = 0, 1$ and Y_i is distributed Bernoulli($p(X_i)$))

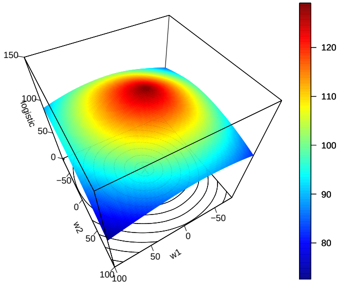
$$
\begin{aligned}
\text{log likelihood}(\beta) &= \log \prod_{i=1}^{n} \left(p(X_i)^{Y_i} (1 - p(X_i))^{(1-Y_i)} \right) \\
&= \sum_{i=1}^{n} (Y_i \log(p(X_i)) + (1 - Y_i) \log(1 - p(X_i)) \\
&= \sum_{i=1}^{n} \Big(Y_i \log(e^{\beta^\top X_i} / (1 + e^{\beta^\top X_i})) \\
&\qquad - (1 - Y_i) \log(1 + e^{\beta^\top X_i}) \Big) \\
&= \sum_{i=1}^{n} \left(Y_i \beta^\top X_i - \log(1 + e^{\beta^\top X_i}) \right)
\end{aligned}
$$

(The book writes $\log \text{likelihood}(\beta)$ as $\log L(p)$ in equation (12.1))

Multiple vs. Logistic regression

The nature of the supervisor is quite different, but observe:





(Note that maximizing (something) is the same as minimizing -(something))

Note $\ell(\beta) = -\log \text{likelihood}(\beta)$ is a loss function

Penalized linear regression

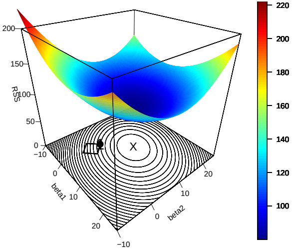
Reminder:

We generalize multiple linear regression to the elastic net by adding a penalty term:

$$SSE_{Enet} = \sum_{i=1}^{n}(Y_i - (\beta_0 + \sum_{j=1}^{p}\beta_j x_j))^2 + \lambda(\alpha\sum_{j=1}^{p}|\beta_j| + (1-\alpha)\sum_{j=1}^{p}\beta_j^2)$$

(Still minimizing this criterion over all β to form the elastic net solution)

For $\alpha = 1$, this looks like $\rightarrow$



The logistic elastic net

Logistic elastic net

We can maximize over β_0, β to find

$$\log L(p) - \lambda(\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2)$$

(Note that this is 'minus' the penalty as it is phrased as a maximization problem)

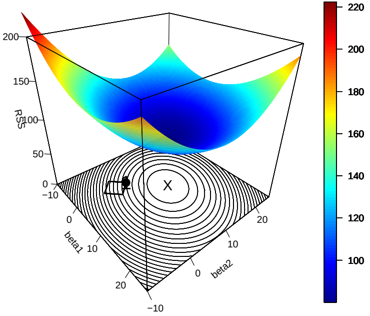
This is the logistic elastic net

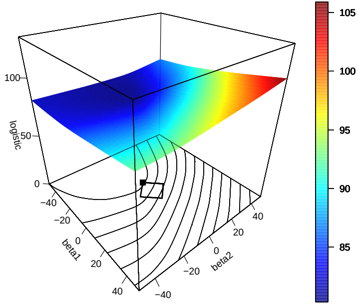
Logistic elastic net

Let's plot the 'minimization' version:

$$-\log L(p) + \lambda(\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2)$$

(The pictures have $\alpha = 1$ and represent the constrained form)





Logistic elastic net in R

We can use the R package glmnet to find the elastic net solution, e.g.:

```
logisticEnet = glmnet(x = X, y = Y, family = 'binomial')
```

Or we can use the familiar caret package function train