Discrete Response Model Lecture 3

datascience@berkeley

Quadratic Term: An Example

Example

Suppose x represents the distance of the placekick. Below is how we can estimate the model.

$$\text{logit}(\pi) = \beta_0 + \beta_1 x + \beta_2 x^2$$

```
mod.fit.distsq<-glm(formula = good ~ distance + I(distance^2), family =
binomial(link = logit), data = placekick)
summary(mod.fit.distsq)
```

Example

```
Call:
glm(formula = good ~ distance + I(distance^2), family = binomial(link = logit),
    data = placekick)

Deviance Residuals:
    Min      1Q    Median      3Q      Max
-2.8625  0.2175  0.2175  0.4011  1.2865

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     7.8446831  1.0009079    7.838 4.59e-15 ***
distance       -0.2407073  0.0579403   -4.154 3.26e-05 ***
I(distance^2)   0.0017536  0.0007927    2.212    0.027 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1013.43  on 1424  degrees of freedom
Residual deviance:  770.95  on 1422  degrees of freedom
AIC: 776.95

Number of Fisher Scoring iterations: 6
```

The estimated models is

$$\text{logit}(\hat{\pi}) = 7.8447 - 0.2407x + 0.001754x^2$$

The p-value for the Wald test of $H_0: \beta_2 = 0$ vs. $H_a: \beta_2 \neq 0$ is 0.027, suggesting there is marginal evidence of a quadratic relationship between distance and the response.

Likelihood Ratio Test

A LRT provides a similar p-value.

```
library(package = car)
Anova(mod.fit.distsq)
```

```
Analysis of Deviance Table (Type II tests)

Response: good
               LR Chisq Df Pr(>Chisq)
distance        16.9246  1  3.889e-05 ***
I(distance^2)    4.7904  1    0.02862 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
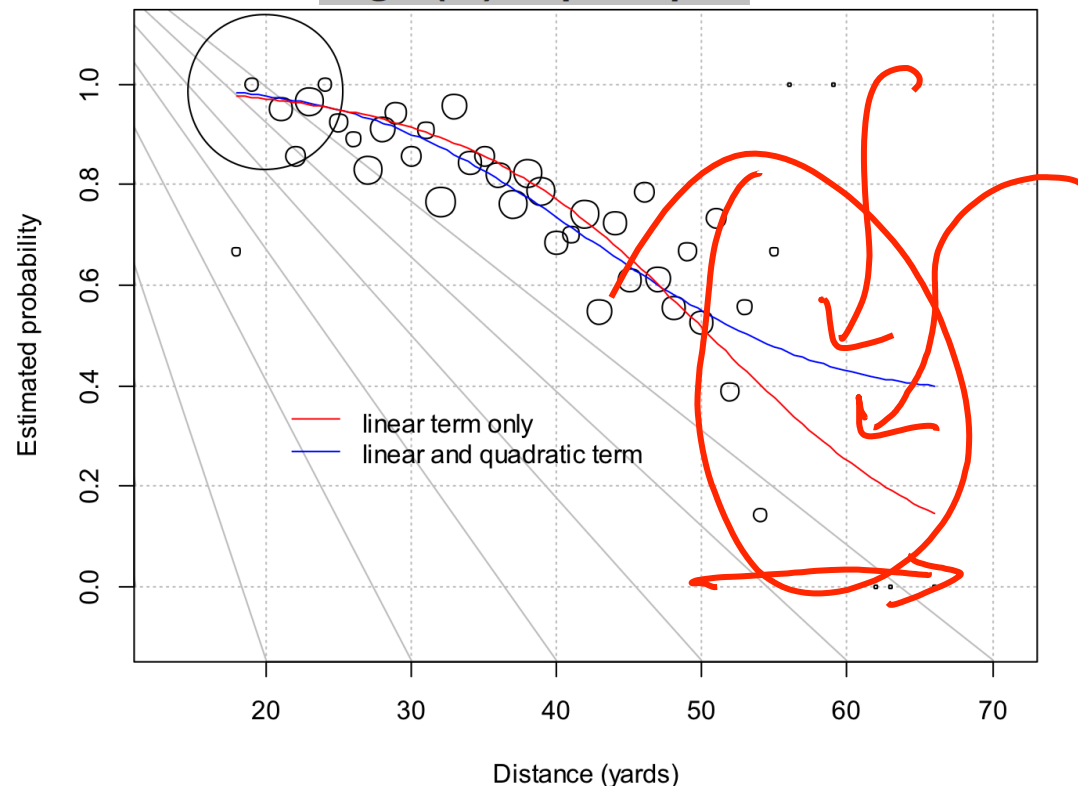
Plot of the Estimated Model

Below is a plot of the estimated model, where the estimate of

$$\text{logit}(\pi) = \beta_0 + \beta_1 x$$

The main difference between the two models appears to be for the larger distances. Given the small number of observations at those distances, it may be difficult to justify the need for the quadratic term.