STATISTICS 538, Lecture #12

From lines to curves

December 1, 2010

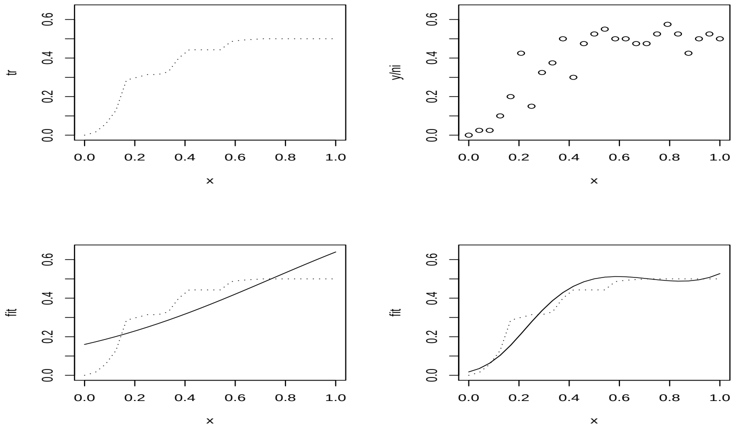
GLM: linear predictor is linear in **parameters**, not necessarily linear in **predictors**

E.g., Y indicates death, X indicates dose of nasty chemical. First thought

$$Pr(Y = 1|X) = \text{expit}(\beta_0 + \beta_1 X)$$

No good? Change link function? Or expand linear predictor from $\eta = \beta_0 + \beta_1 X$ to something more complex (but still linear in parameters).

Consider this true relationship spawning these data, and two fits



Alternative to polynomials? Splines!

Cubic splines: Choose **knots** $c_1, \ldots, c_q$. Consider piecewise cubic functions (between knots) with smooth joins (at knots)

Splines have 'basis function' representations, i.e., linear in parameters

So 'cheap homemade' spline fitting...

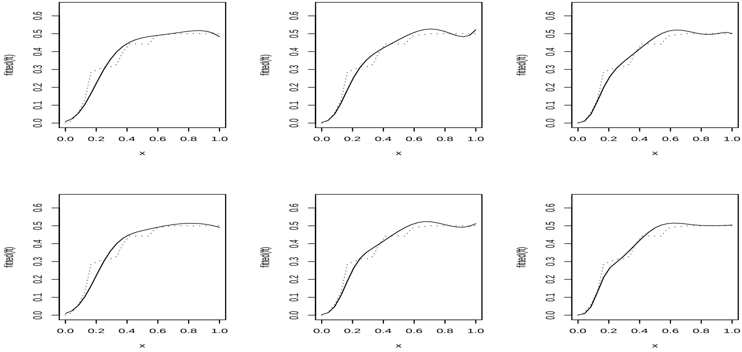
```
xmat <- cbind(x,x^2,x^3)

for (j in 1:q) {
  xmat <- cbind(xmat, pmax(x - quantile(x,j/(q+1)),0)^3)
}

ft <- glm(cbind(y,ni-y) ~ xmat, family=binomial)

plot(x, fitted(ft)) ...
```

Add one, two, three parameters beyond the cubic model: polynomials (top) versus splines (bottom)



Further issues

Model selection: how many knots?

More refined splines: **natural cubic splines**

Further issues, continued

Lots of curve-fitting packages - don't need to do-it-yourself

```
> library(splines)
> ft <- glm(cbind(y,ni-y)~ ns(x,df=4), family=binomial)
```