Class Notes

Statistical Computing & Machine Learning

Class 18

Smoothers

Sometimes the model form we want to impose is described by broad properties:

1. The "smoothness" of the model.
2. The generalizability of the model, e.g. extrapolation outside the range of the inputs

The broad class of model forms used here are called smoothers. They are a linear combination of a set of functions, called *basis functions*, that have nice properties.

Ideas of smoothness

- Continuity: Is the nth derivative continuous? The higher is n, the smoother is the function.
- Bumpiness: The integral of the absolute value of the 2nd derivative.

Polynomials

Polynomials have been central to math education for a long time, and there has been a rich theory of them since around the 13th century. For instance:

> The fundamental theorem of algebra states that every non-constant single-variable polynomial with complex coefficients has at least one complex root. This includes polynomials with real coefficients, since every real number is a complex number with an imaginary part equal to zero.

- Polynomials are completely smooth in the sense of continuity: all derivatives are continuous.
- But they may be bumpy
- And their behavior for large and small x is crazy.

The model matrix

The set of basis functions evaluated at the inputs x.

```
make_model_matrix <- function(x, basis_funs) {
  MM <- matrix(0, nrow=length(x), ncol=length(basis_funs))
  for (i in 1:length(basis_funs)) {
    MM[,i] = basis_funs[[i]](x)
  }

  return(MM)
}
```

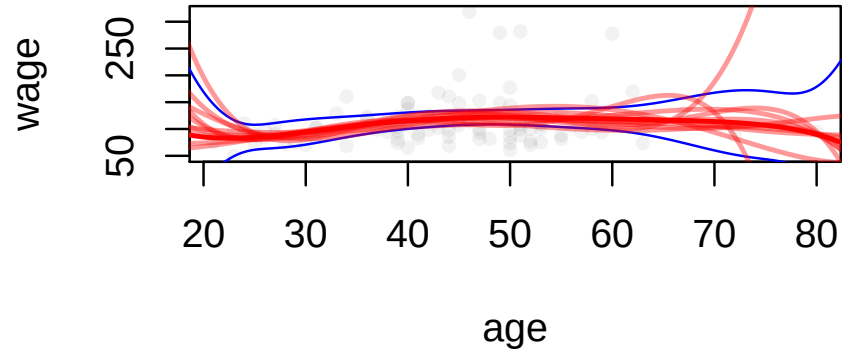
Polynomial basis functions:

```
polynomial_basis_3 <- list(
  function(x) 1,
  function(x) x,
  function(x) x^2,
  function(x) x^3
)

monomial <- function(k) function(x) x^k

make_polynomial_basis <- function(p) {
  lapply(0:p, FUN=monomial)
}

show_smoother(basis=make_polynomial_basis(5), data=mosaic::sample(Wage, size=100), bootstrap=10, confidenc
```



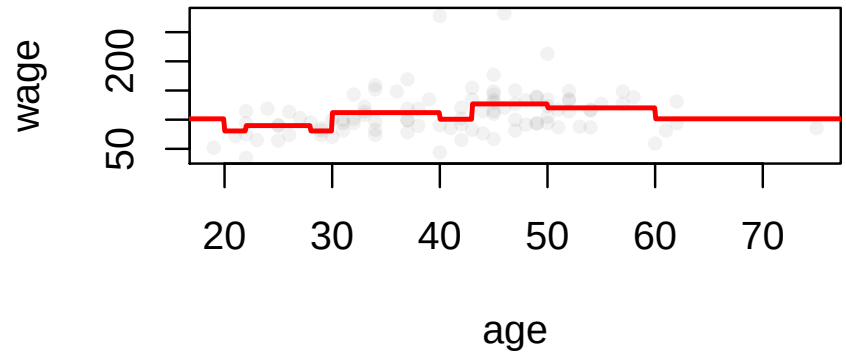
Sigmoidal Functions

```
sigmoidal_basis <- list(
  function(x) 1,
  function(x) dnorm(x, mean=25, sd = 10),
  function(x) dnorm(x, mean=40, sd = 15),
  function(x) dnorm(x, mean=55, sd=10)
)
```

Hat functions

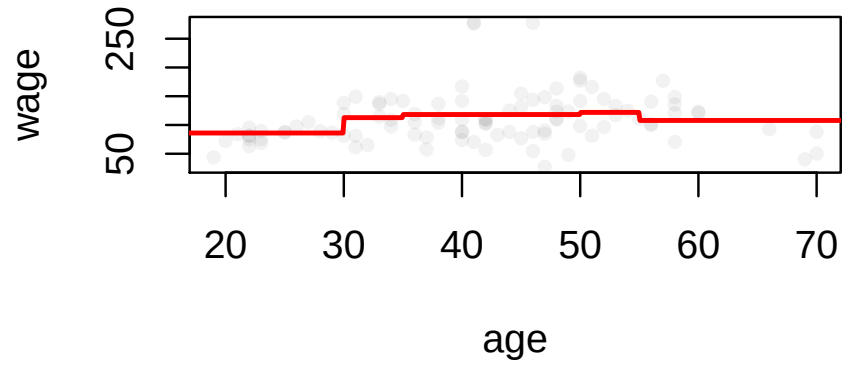
```
hat <- function(from, width)
  function(x) {ifelse(x>from & x<(from+width), 1, 0)}
hat_basis <- list(
  function(x) 1,
  hat(20, 10),
  hat(30, 10),
  hat(40, 10),
  hat(43, 17),
  hat(22, 6),
  hat(50, 10)
)

show_smoother(basis=hat_basis, data=mosaic::sample(Wage, size=100), bootstrap=0, confidence=FALSE)
```



Steps

```
step_fun <- function(where) { function(x) ifelse(x > where, 1, 0)}
step_basis <- list(
  function(x) 1,
  step_fun(30),
  step_fun(35),
  step_fun(50),
  step_fun(55)
)
show_smoother(basis=step_basis, data=mosaic::sample(Wage, size=100), bootstrap=0, confidence=FALSE)
```



Other functions

- triangles
- gaussian: `dnorm()`
- sigmoidals: `pnorm()`
- spline basis: 1, x, x^2, x^3, $(x-\xi_j)^3_+$

Holes in the data

Leave out the middle of the data

Bootstrapping

Normal theory confidence bands

- covariance matrix for model coefficients
- `rowSums(MM %*% cov * MM)`

Smoothers in k dimensions

Programming Activity

Day 16 Programming Activity. Generating data and fitting models.