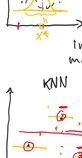
Week 6

Monday, 15 February 2021 12:10

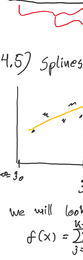
4.3)

Parameter h in local regression

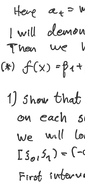


Should h be estimated by ML?

No, because it will yield a small h which in turn will overfit the mode to the training data.



4.5) Splines



We will look at eq 4.10.

$$f(x) = \sum_{j=1}^{K+4} \beta_j h_j(x),$$

where $h_j = x^{j-1}$, for $j = 1,2,3,4$.

and $h_{j+4} = (x - \xi_j)_+^3$ for $j = 1, 2, \dots, K$.

(number of knots)

Here $a_+ = \max(0, a)$

I will demonstrate it for $K=1$.

Then we have that

(*) $f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + \beta_5 (x - \xi_1)_+^3$

1) Show that f is a cubic function on each subinterval $[\xi_j, \xi_{j+1})$ for $j = 1, \dots, K+1$

We will look at two intervals $[\xi_0, \xi_1) = (-\infty, \xi_1)$ and $[\xi_1, \xi_2) = [\xi_1, \infty)$.

First interval: $x < \xi_1$, so

$f_- = f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3$,

which is a cubic function.

Second interval: $x \geq \xi_1$

$f_+ = f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + \beta_5 (x - \xi_1)^3$

$= \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 - \beta_5 x^3 + 3\beta_5 x^2 \xi_1 - 3\beta_5 x \xi_1^2 + \beta_5 \xi_1^3$

$= (\beta_1 - \beta_5 \xi_1^3) + (\beta_2 + 3\beta_5 \xi_1^2) x + (\beta_3 - 3\beta_5 \xi_1) x^2$

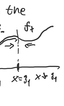
$+ (\beta_4 + \beta_5) x^3$.

This is also a cubic function.

2) Show that f has two continuous derivatives.

We only need to look at the knot value ξ_1.

Compute the 1st and 2nd derivative of f_- and f_+.



$f_-'(x) = \beta_2 + 2\beta_3 x + 3\beta_4 x^2$

$f_-''(x) = 2\beta_3 + 6\beta_4 x$

$f_+'(x) = \beta_2 + 2\beta_3 x + 3\beta_4 x^2 + 3\beta_5 (x - \xi_1)^2$

$= \beta_2 + 2\beta_3 x + 3\beta_4 x^2 + 3\beta_5 x^2 - 6\beta_5 x \xi_1 + 3\beta_5 \xi_1^2$

$f_+''(x) = 2(\beta_3 - 3\beta_5 \xi_1) + 6(\beta_4 + \beta_5) x$

$= 2\beta_3 + 6\beta_4 x + 6\beta_5 x - 6\beta_5 \xi_1$

Then look at the limits

$\lim_{x \uparrow \xi_1} f_-''(x) = \lim_{x \uparrow \xi_1} 2\beta_3 + 6\beta_4 x = 2\beta_3 + 6\beta_4 \xi_1$

The other limit

$\lim_{x \downarrow \xi_1} f_+''(x) = \lim_{x \downarrow \xi_1} 2\beta_3 + 6\beta_4 x + \underbrace{6\beta_5 (x - \xi_1)}_{=0}$

$= 2\beta_3 + 6\beta_4 \xi_1$

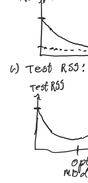
This means that f has two cont. derv.

3) We will see that we get a jump in the third derivative at ξ_1.

Just differentiate f_-'' and f_+'' once more, and we get

$f_-'''(x) = 6\beta_4$ and $f_+'''(x) = 6\beta_4 + 6\beta_5$

we see that the jump is $6\beta_5$.



4.6)

Prove eq 4.12.

We want to show that the $\hat{\theta}$ which minimizes $D = (y - N\theta)^T (y - N\theta) + \lambda \theta^T \Omega \theta$ is given by $\hat{\theta} = (N^T N + \lambda \Omega)^{-1} N^T y$.

And Ω is a penalty matrix.

(symmetric)

Rewriting D

$D = y^T y - y^T N\theta - \theta^T N^T y + \theta^T N^T N \theta + \lambda \theta^T \Omega \theta$

$= y^T y - (N^T y)^T \theta - \theta^T (N^T y) + \theta^T (N^T N) \theta + \lambda \theta^T \Omega \theta$

I use the matrix cookbook and eq 69 and 81 to do the derivatives.

$\frac{\partial D}{\partial \theta} = -N^T y - N^T y + [N^T N + (N^T N)^T] \theta + \lambda(\Omega + \Omega^T)\theta$

$= -2N^T y + 2N^T N\theta + 2\lambda\Omega\theta$

Set it equal to zero and solve for θ.

$0 = -2N^T y + 2N^T N\theta + 2\lambda\Omega\theta \quad \left| \cdot \frac{1}{2} \right.$

$N^T y = (N^T N + \Omega)\theta \quad \left| \cdot (N^T N + \lambda\Omega)^{-1} \right.$

$\hat{\theta} = (N^T N + \lambda\Omega)^{-1} N^T y.$

∎

ISLR - exercises.

6.2.

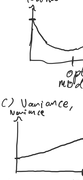
a) Lasso: solution iii.

b) Ridge: solution ii)

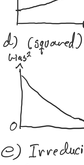
c) non-linear methods: solution ii)

6.3) s down → 0 s up → ∞ Lasso

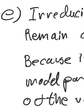
a) Traing RSS: iv)



b) Test RSS: ii)



c) Variance, iii)



d) (squared) bias: iv)

e) Irreducible error, v)

Remain constant

Because it is independent of model parameters and thus independent of the value of s.