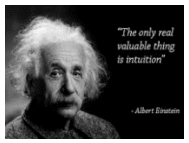
Smoothing Splines

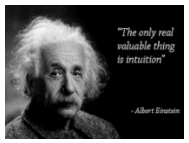


Smoothing Splines: Intuition

- The best way to think about a smoothing spline is to think about a spline with some **rounding** around the **knots** – i.e., smooth knots.
- Regular **splines** are constructed by using **dummy variables** to differentiate data before and after the knots, and **truncated power functions** to fit the model in the particular model section, and then estimating the resulting model using **OLS**
- If you take this spline model and **estimate** it using a **penalized method** similar to **Ridge** regression, the result is a **smoothing spline**
- The end result is a spline model but with its **"roughness" smoothed** out by shrinking (i.e. penalizing) coefficients where there is more variability (i.e., roughness) in the curve, **particularly** around the **knots**.
- The **degree** of **"smoothing"** can be **controlled** with a **tuning parameter** λ similar to Ridge regression

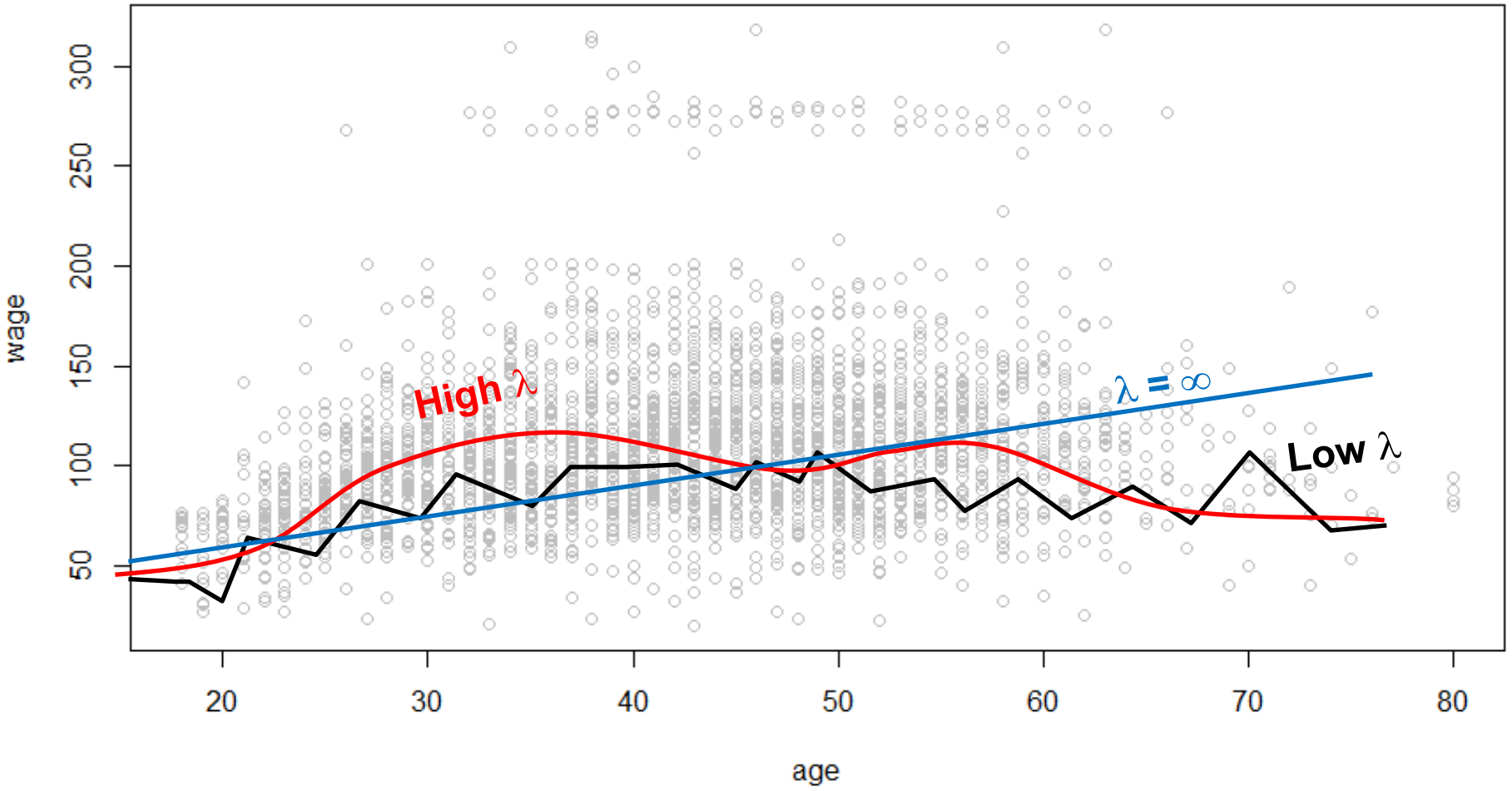
Smoothing Splines Explained

- λ in a smoothing spline is similar to the λ in Ridge regression – i.e., it **penalizes variability**
- But in soothing splines, the **penalty** is applied on the **variability** (i.e., roughness or second derivative) of the curve
- For example, a **straight line** has a **constant slope**, whereas the **slope** in a **curve** changes in **every point** of the curve.
- λ penalizes this variability in slope more strongly where the slope changes more strongly
- And the **larger** the λ the smoother the curve →
  - $\lambda=0$ produces a jumpy curve that touches every training data point
  - As λ **increases** the curve becomes **smoother** → equivalent to a **piecewise cubic** with a **knot** in **every data point** → the notion of **"knots"** in smoothing splines becomes **irrelevant**
  - When $\lambda = \infty$ the curve is perfectly smooth – i.e., a **straight line** fitted with **OLS**
- As with other models, the **optimal** λ is selected with **cross-validation**



Smoothing Spline Illustration

The higher the λ the smoother the smoothing spline line



Tips

`smooth.spline(){stats}` → "Smooth Spline" function in the `{stats}` package to fit smooth spline models

`fit.smooth1=smooth.spline(age,wage,df=16)` → Fits a smooth spline model with an arbitrary `df=16`; this function finds the lambda and CV corresponding to `df=16`

`fit.smooth2=smooth.spline(age,wage,cv=TRUE)` → This model lets cross-validation find:

The best lambda → `fit.smooth2$lambda`

The corresponding df → `fit.smooth2$df`

KOGOD SCHOOL
of
BUSINESS