Stat 432 Homework 9

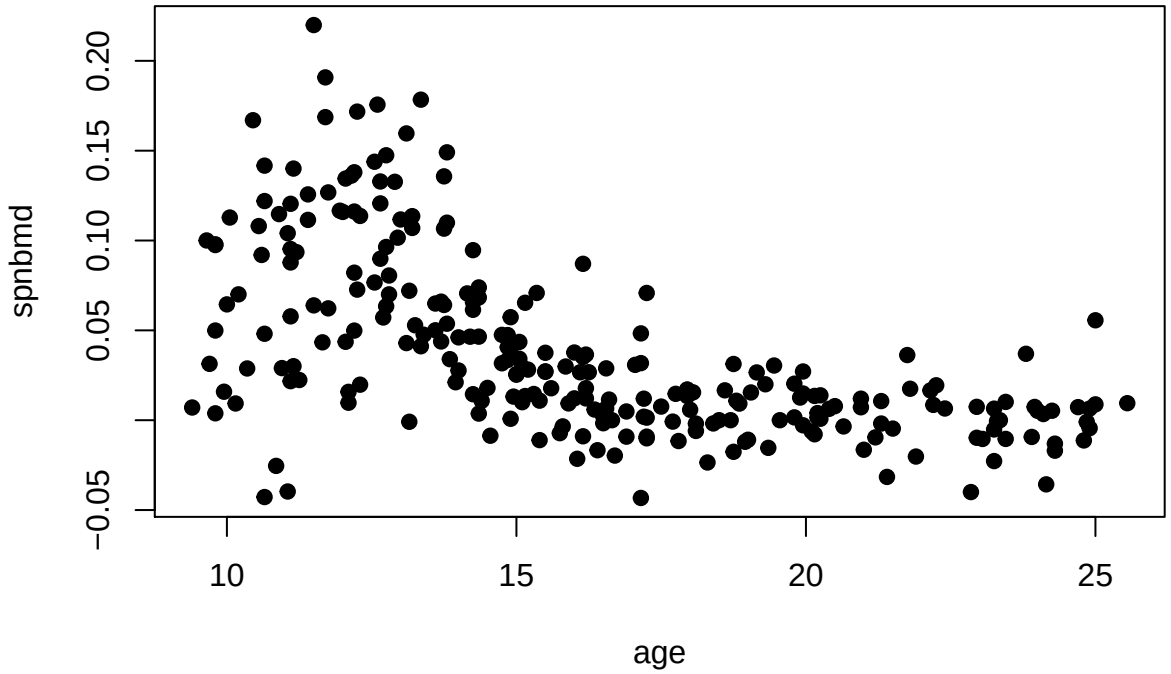
Assigned: Mar 30, 2019; Due: 11:59PM Apr 5, 2019

Before starting this homework, you should read the rlab file of Spline on the course website carefully.

Question 1 (Cubic Spline) [10 points]

We demonstrated how to construct the spline basis using $(\cdot)_+$ functions to force continuity. If we are interested in having a higher order of smoothness, we can consider increasing the power of our spline basis. Following this idea, use the definitions on pages 23 and 32 of the lecture note, perform the following question using the female portion of the `bone` dataset.

```
library(ElemStatLearn)
data(bone)
traindata = bone[bone$gender == "female", ]
plot(spnbmd~ age, data = traindata, pch = 19)
```



- Pick 2 cut points as knots, based on your preference.
- Construct the cubic spline basis by writing your own code. What is the degree of freedom?
- Fit linear regression (you can use `lm()`) using these basis functions and plot the fitted values.
- Using the `bs()` function, in combination with `lm()` to fit the exact same linear regression. Demonstrate your result using plots.
- Construct the natural cubic spline basis by writing your own code. What is the degree of freedom?
- Repeat what you did for the cubic spline basis.

Question 2 (Multiple Variables in Spline) [extra-credit 3 points]

We demonstrated that fitting multiple variables of spline can be done using an additive structure. Read the example in the rlab file of `spline`, and perform the following.

```
data(ozone)
head(ozone)
```

```
##   ozone radiation temperature wind
```

```
## 1    41       190          67  7.4
## 2    36       118          72  8.0
## 3    12       149          74 12.6
## 4    18       313          62 11.5
## 5    23       299          65  8.6
## 6    19        99          59 13.8
```

The ozone data is trying to model the ozone level using other 3 variables: `radiation`, `temperature` and `wind`.

- Using the functions you developed in Question 1, fit this multivariate spline model using an additive structure. You should be using at least 1 knot for each variable, other than that, the choice is yours. You do not need to tune your method, but feel free to do so.
- Compare your result with an additive model using the built-in Natural Cubic Splines basis (the `ns()` function). You need to construct the NCS basis such that the number of degrees of freedom matches exactly your own construction in the first part. Other than that, the choice is yours.
- Comment on the difference between these two results, no matter which is better.