

Splines and over-fitting

2023-11-14

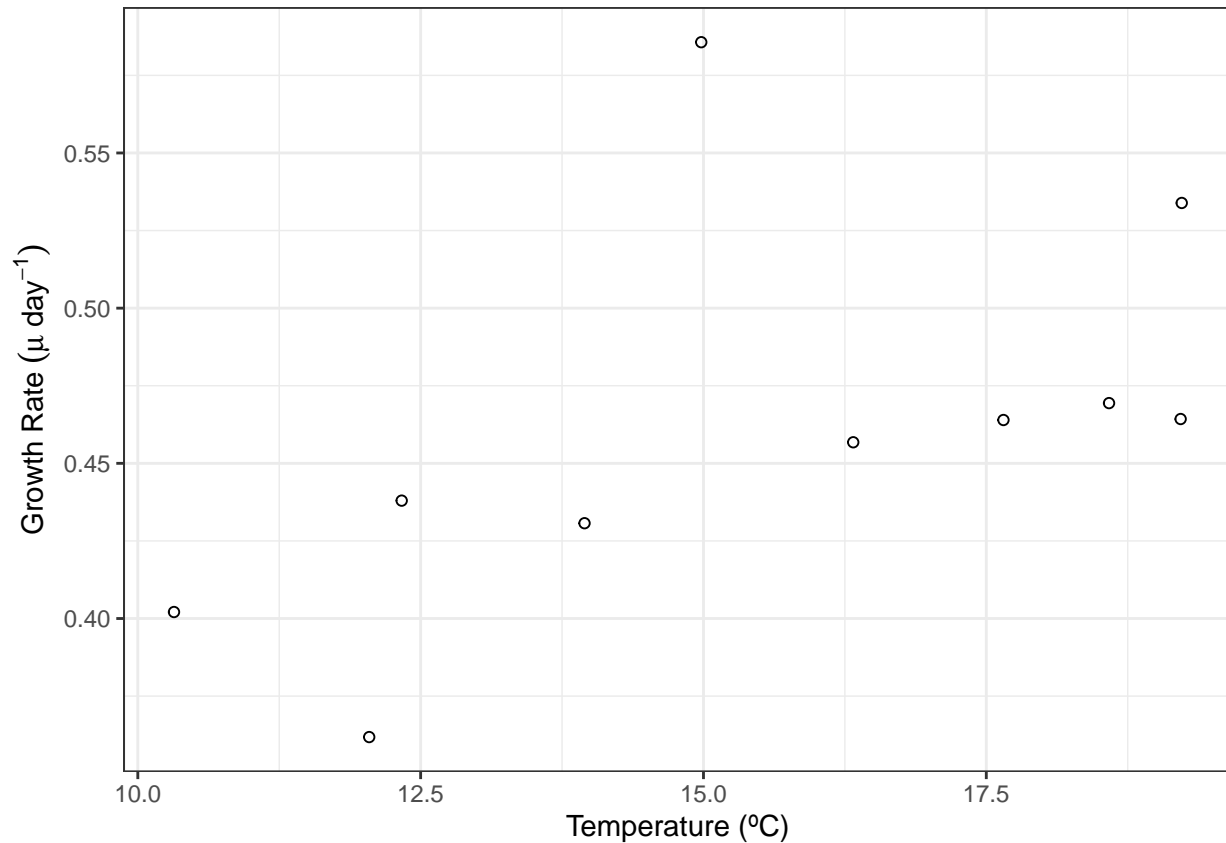
R Markdown

Cell growth rate (per day) in diatom of genus *Thalassiosira* taken from Montagnes and Franklin 2001.

```
f_temp <- function(sst){rlnorm(1,meanlog=log(0.014*sst + 0.25),sdlog = 0.1)}  
set.seed(2003)  
  
x_temp <- runif(30,min=10,max=20)  
y_growth_rate <- vapply(x_temp,f_temp,FUN.VALUE = 0)  
  
df_growth <- data.frame(temp=x_temp,growth_rate=y_growth_rate)
```

Including Plots

Let's just consider the first 10 data points and assume that's our initial data set.



Fit a linear regression.

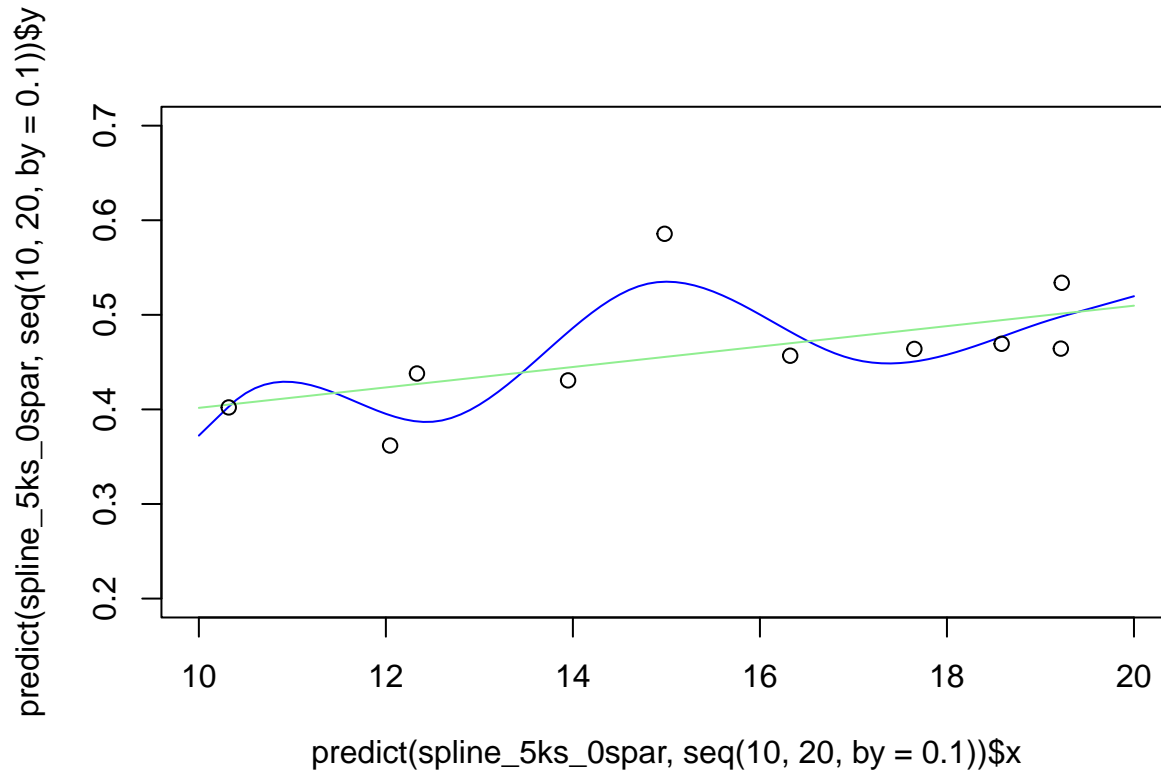
Fit a spline.

```
?smooth.spline
```

There's a lot of ways R lets you control how this procedure works; to be honest I'm not clear on how all these operate. The original notions of splines worked on "knots", though.

```
spline_5ks_0spar <- smooth.spline(x_temp[1:10], y_growth_rate[1:10], all.knots = seq(0,1,by=.25), spar=0.0)
spline_5ks_1spar <- smooth.spline(x_temp[1:10], y_growth_rate[1:10], all.knots = seq(0,1,by=.25), spar=1.0)

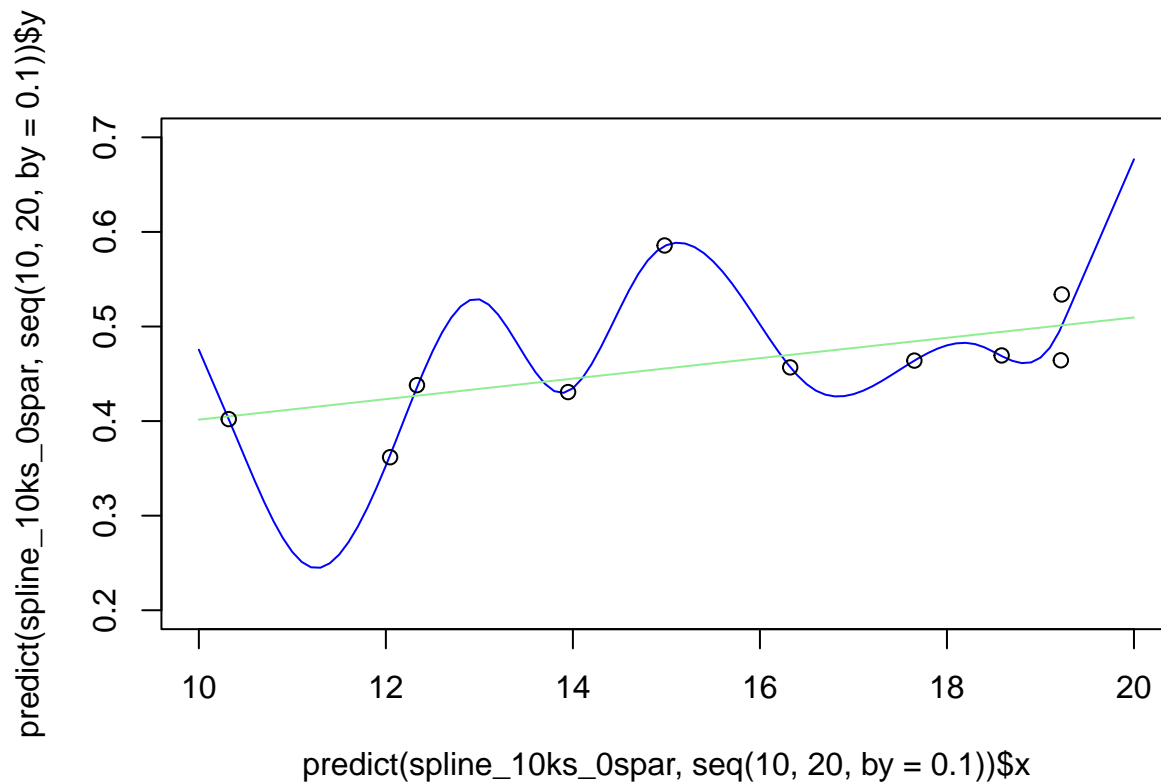
plot(predict(spline_5ks_0spar, seq(10,20,by=0.1)), type="l", col="blue", ylim=c(0.2,.7)) +
  lines(predict(spline_5ks_1spar, seq(10,20,by=0.1)), col="lightgreen") +
  points(x_temp[1:10], y_growth_rate[1:10])
```



```
## integer(0)
```

```
spline_10ks_0spar <- smooth.spline(x_temp[1:10], y_growth_rate[1:10], all.knots = seq(0,1,by=.1), spar=0)
spline_10ks_1spar <- smooth.spline(x_temp[1:10], y_growth_rate[1:10], all.knots = seq(0,1,by=.1), spar=1)

plot(predict(spline_10ks_0spar, seq(10,20,by=0.1)), type="l", col="blue", ylim=c(0.2,.7)) +
  lines(predict(spline_10ks_1spar, seq(10,20,by=0.1)), col="lightgreen") +
  points(x_temp[1:10], y_growth_rate[1:10])
```



```
## integer(0)
spline_auto <- smooth.spline(x_temp[1:10], y_growth_rate[1:10])
print(spline_auto)

## Call:
## smooth.spline(x = x_temp[1:10], y = y_growth_rate[1:10])
##
## Smoothing Parameter spar= 1.245035 lambda= 813.5593 (30 iterations)
## Equivalent Degrees of Freedom (Df): 1.99979
## Penalized Criterion (RSS): 0.02475346
## GCV: 0.003867525

plot(predict(spline_auto, seq(10, 20, by = 0.1)), type = "l", col = "lightgreen")
abline(a = 0.25, b = 0.014)
```

