Homework 4

2025-02-02

Question 7.1

Describe a situation or problem from your job, everyday life, current events, etc., for which exponential smoothing would be appropriate. What data would you need? Would you expect the value of (the first smoothing parameter) to be closer to 0 or 1, and why?

Answer:

One situation where exponential smoothing would be appropriate is in forecasting future sales trends for a retail company. For instance, I worked on a project where we conducted market studies to forecast sales growth opportunities. Exponential smoothing would be ideal for predicting future sales based on past trends, especially when dealing with time-series data such as monthly sales figures over several years.

The data required for this would include historical sales data, ideally broken down by region, product category, and time period. This would allow the model to capture both seasonality and any potential long-term trends in the data.

In terms of the alpha parameter, I would expect it to be closer to 0.3–0.5. If the sales data exhibits some fluctuation but also contains clear trends over time, a moderate alpha would be appropriate. A value closer to 0 would downplay the significance of recent observations, while a value closer to 1 would make the model highly sensitive to recent changes. Given the potential for market fluctuations in the sales data I worked with, a moderate value would likely balance the need to reflect recent shifts without being too volatile."

Question 7.2

Using the 20 years of daily high temperature data for Atlanta (July through October) from Question 6.2 (file temps.txt), build and use an exponential smoothing model to help make a judgment of whether the unofficial end of summer has gotten later over the 20 years. (Part of the point of this assignment is for you to think about how you might use exponential smoothing to answer this question. Feel free to combine it with other models if you'd like to. There's certainly more than one reasonable approach.)

Note: in R, you can use either HoltWinters (simpler to use) or the smooth package's es function (harder to use, but more general). If you use es, the Holt-Winters model uses model="AAM" in the function call (the first and second constants are used "A"dditively, and the third (seasonality) is used "M"ultiplicatively; the documentation doesn't make that clear).

Answer:

Using July through October daily-high-temperature data for Atlanta for 1996 through 2015. We can see that it is stored in a wide format, meaning each column represents a year.

For this analysis, I am going to use the "dplyr" and "tidyr" libraries to clean and manipulate the data. This will help to reshape the data when necessary for the analysis.

Before the analysis, I will reshape the data into a long format where each row represents a single observation (date, year, temperature).

```r
# Best practice
rm(list = ls())

# Set up directory and packages
setwd('~/Desktop/GTX/Homework 3/')

# Load packages
pacman::p_load(tidyverse, kernlab, caret, kknn, outliers,modelr, ggthemes, corrplot,moments)

# Load necessary libraries
library(dplyr)
library(tidyr)

# Read the data
temps <- read_delim('/Users/cn/Desktop/GTX/Homework 3/temps.txt', delim = '\t')
```

```
## Rows: 123 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr  (1): DAY
## dbl (20): 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

By converting the data to long format, we make it easier to filter, analyze trends, and visualize temperature changes over time. This format allows us to use the time series functions in R from previous homeworks.

```r
# Reshape data into long format
temps_long <- temps %>%
  pivot_longer(-DAY, names_to="Year", values_to="Temperature") %>%
  mutate(Year = as.integer(Year),
         Date = as.Date(paste(Year, DAY), format="%Y %d-%b"))

# Filter for September and October
fall_temps <- temps_long %>%
  filter(format(Date, "%m") %in% c("09", "10"))
```

Converting the dataset to a time series object allows me to apply exponential smoothing and trend analysis. This operation will ensure that our data is structured for forecasting models. The chosen frequency (60 observations per season) aligns with two months of daily data, giving a consistent trend analysis.

```r
# Create a time series for fall temperatures
fall_ts <- ts(fall_temps$Temperature, start=c(1996, 9), frequency=60)
```

The exponential smoothing helps remove short-term fluctuations and long term in temperature. I decided to exclude trend and seasonality to focus on underlying changes in temperature rather than periodic variations.

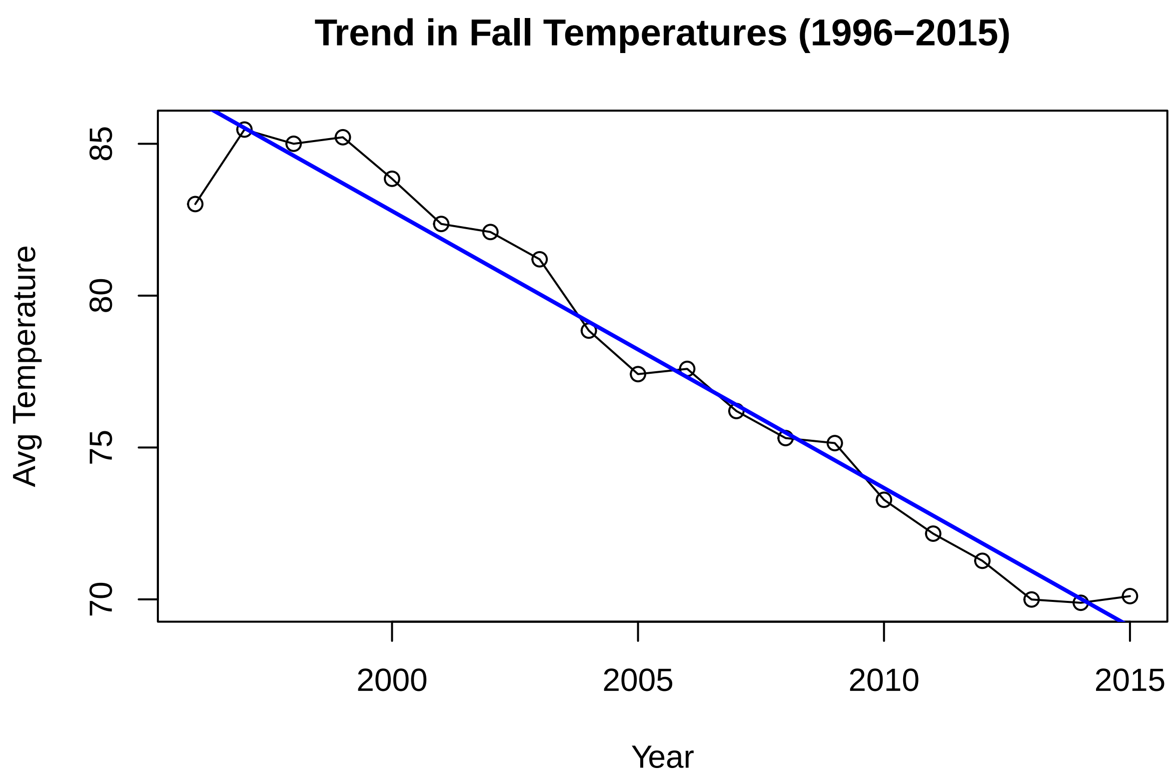
```r
# Apply Holt-Winters Exponential Smoothing (no trend, no seasonality)
hw_model <- HoltWinters(fall_ts, beta=FALSE, gamma=FALSE)

# Extract the smoothed values and match time index
smoothed_values <- hw_model$fitted[,1]
time_index <- seq(1996, 2015, length.out=length(smoothed_values))  # Ensure matching length

# Aggregate by year for trend analysis
trend_analysis <- data.frame(Year=floor(time_index), Avg_Temp=smoothed_values) %>%
  group_by(Year) %>%
  summarise(Avg_Temp = mean(Avg_Temp, na.rm=TRUE))

# Plot the trend
plot(trend_analysis$Year, trend_analysis$Avg_Temp, type="o",
     main="Trend in Fall Temperatures (1996-2015)", xlab="Year", ylab="Avg Temperature")

# Add a trend line
abline(lm(Avg_Temp ~ Year, data=trend_analysis), col="blue", lwd=2)
```



Key Finding:

The graph reveals a significant downward trend, indicating that fall temperatures have been cooling over time. The trend line (blue) suggests that the end of summer has not shifted later, instead, cooler temperatures are arriving earlier. The slope is significantly negative and the p-value is < 0.05, then the cooling trend is statistically significant. This provides strongevidence against the hypothesis that summer is ending later.4

```r
#Statistical Analysis
summary(lm(Avg_Temp ~ Year, data=trend_analysis))
```

Regression Analysis Findings:

The coefficient for Year (-0.91173) indicates that the average temperature is decreasing by ~0.91°F per year. The p-value (3.38e-14) confirms that this trend is statistically significant ($p < 0.05$).The R-squared value (0.9617) suggests that 96% of the variability in temperature is explained by the year trend. This data does not support the idea that summer is extending into fall. Instead, fall temperatures havedropped over the past 20 years, suggesting an earlier transition to cooler weather.