DATA 1010
SAMUEL S. WATSON
3 DECEMBER 2018

There are no class problem sets for the days from November 16 to December 3.

On November 16, we continued working through the neural nets material (Section 2.4 in the text, excluding Section 2.4.1).

On November 19, we had two two-hour sessions. We finished the neural nets material, and then we tackled dimension reduction (Section 2.5: PCA and t-SNE).

On November 26, we talked about the likelihood ratio classifier and the Neyman-Pearson lemma (Section 1.2.1). We also discussed neural nets for classification (Section 2.4.1).

On November 28, we started learning the basics of R and `ggplot2`. For the basics of R, we followed the Julia-Python-R cheatsheet. Of particular note are the following ideas: (i) **`sapply`** may be used as a substitute for array comprehensions, and subset selection is performed by producing a boolean vector and indexing with it. For example, `A[A > 0]` forms the vector `A > 0` of **`TRUE`**s and **`FALSE`**s indicating where `A` has positive entries, and indexing with it returns the vector of `A`'s positive entries. For ggplot2, we followed the data visualization section in *R for Data Science*:

`https://r4ds.had.co.nz/data-visualisation.html`

On November 30, we continued in the *R for Data Science* book and covered the `dplyr` section

`https://r4ds.had.co.nz/transform.html`

On December 3, we reviewed for the exam and began discussing point estimation. We learned the terms *statistical functional, estimator, bias, consistency, and mean squared error*.

On December 5, we continued our discussion of point estimation. See Section 4.1 in the second volume of the course text.

On December 7, we learned about confidence intervals, the Glivenko-Cantelli theorem, and the DKW inequality (Section 4.2, 4.3)

On December 10, we reviewed the DKW inequality and covered bootstrapping, introduced maximum likelihood estimation, and discussed hypothesis testing. See Sections 4.4 through 4.6 in the course text.