David J. Olive

Prediction and Statistical Learning

July 2, 2023

Preface

Many statistics departments offer a one semester graduate course in statistical learning theory using texts such as Hastie et al. (2009), Hastie et al. (2015), and James et al. (2021). Also see Berk (2016), Izenman (2008), Kuhn and Johnson (2013), Marden (2006), and Murphy (2012).

The prerequisite for this text is a calculus based course in statistics at the level of Chihara and Hesterberg (2011), Hogg, Tanis, and Zimmerman (2020), Larsen and Marx (2011), Wackerly, Mendenhall and Scheaffer (2008) or Walpole, Myers, Myers and Ye (2016). Linear algebra and one computer programming class are essential. Knowledge of regression would be useful. See Olive (2017a) and Cook and Weisberg (1999). Knowledge of multivariate analysis would be useful. See Olive (2017b) and Johnson and Wichern (2007).

Some highlights of this text follow.

- Prediction intervals are given that can be useful even if $n < p$.
- The response plot is useful for checking the model.
- The large sample theory for the elastic net, lasso, and ridge regression is greatly simplified.
- The large sample theory for some data splitting estimators, variable selection estimators, marginal maximum likelihood estimators, and one component partial least squares will be given. See Olive and Zhang (2023).

Downloading the book's R functions *slpack.txt* and data files *sldata.txt* into *R*: The commands

```
source("http://parker.ad.siu.edu/Olive/slpack.txt")
source("http://parker.ad.siu.edu/Olive/sldata.txt")
```

The *R* software is used in this text. See R Core Team (2020). Some packages used in the text include `glmnet` Friedman et al. (2015), `leaps` Lumley (2009), `MASS` Venables and Ripley (2010), and `pls` Mevik et al. (2015).

Acknowledgements

Teaching the material to Math 583 students at Southern Illinois University in 2017 was very useful. The text was used for a high dimensional statistics course in 2023. Trevor Hastie's website had a lot of useful information. Work by R. Dennis Cook and his coauthors was useful for figuring out OPLS.

Contents