Class Notes

Statistical Computing & Machine Learning

Class 15

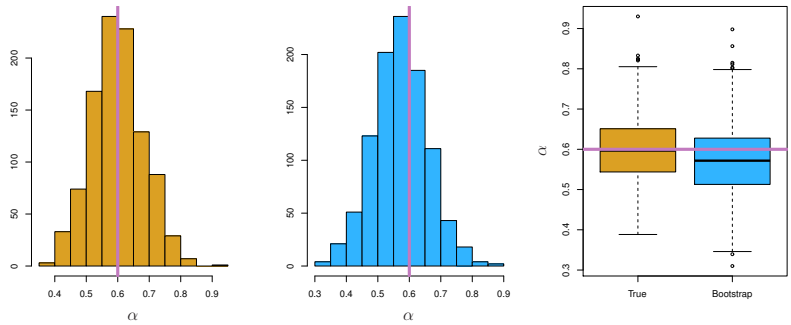
Programming

Functionals. How `optim()` allows you to pass the various other arguments, e.g. `data=` to the function being optimized.

The meaning of `...`

Solvers

Bootstrapping



ISLR Figure 5.10

In-class demonstration. How many cases don't get used in a typical resample?

```
cases <- 1:10000
ntrials <- 50
result <- numeric(ntrials)
for (k in 1:50) {
  bootstrap_sample <- sample(cases, size=length(cases), replace = TRUE)
  result[k] <- length(setdiff(cases, unique(bootstrap_sample))) / length(cases)
}
result
```

```
##  [1] 0.3678 0.3680 0.3675 0.3714 0.3649
##  [6] 0.3615 0.3684 0.3686 0.3613 0.3669
## [11] 0.3648 0.3684 0.3705 0.3651 0.3687
## [16] 0.3694 0.3628 0.3674 0.3673 0.3647
## [21] 0.3694 0.3720 0.3681 0.3719 0.3622
## [26] 0.3666 0.3706 0.3648 0.3722 0.3737
## [31] 0.3700 0.3668 0.3666 0.3691 0.3696
## [36] 0.3666 0.3641 0.3700 0.3663 0.3715
```

```
## [41] 0.3646 0.3692 0.3662 0.3627 0.3712
## [46] 0.3656 0.3677 0.3639 0.3672 0.3689

## Or ...
(1 - 1/length(cases))^length(cases)

## [1] 0.367861
```

Programming a bootstrap calculation

```
bootstrap <- function(formula, method, data, statistic, reps=10) {



}
```

Mosaic software
Using `resample()` and `do()`

Model Selection

Problem: Given a set of potential model terms, choose the best subset.

Two issues:

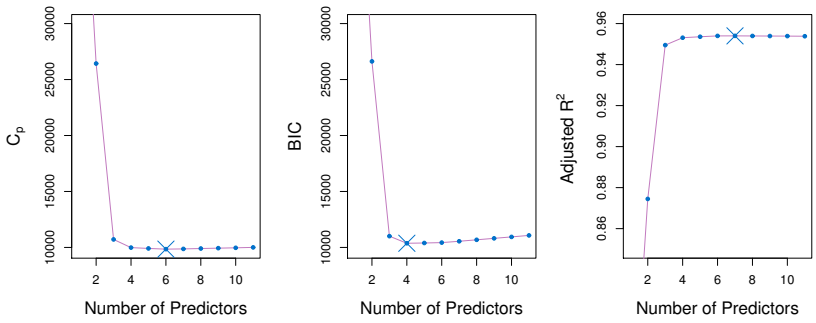
1. What does "best" mean?
2. How to optimize?

Best

In-sample	Adjusted	Out-of-sample
$\frac{1}{n}$RSS	$C_p = \frac{1}{n}(\text{RSS} + 2d\hat{\sigma}^2)$	cross-validated prediction error
.	$\text{AIC} = -2\ln\mathcal{L} - 2d$	.
.	$\text{AIC}_{ls} = \frac{C_p}{\hat{\sigma}^2}$	.
$\text{BIC} = \frac{1}{n}(\text{RSS} + \ln(n)d\hat{\sigma}^2)$	.	
R^2	Adjusted R^2	???

Adjusted $R^2 = 1 - \frac{\text{RSS}/(n-d-1)}{\text{TSS}(n-1)}$

ISLR Figure 6.2. Note that the values on the vertical axis are the best for that "number of predictors."



Uncertainty

Repeat the analysis for different test sets or using different folds in k-fold cross validation.

- At each value of "Number of Predictors", there will be a distribution.
- *One-standard-error rule*: select the smallest model for which the estimated test error is within one standard error of the lowest point on the curve.

Daily activity

Write a k-fold cross validator. Day-13-Programming-Tasks