CASI 7.3 and Ch 16 Highlights

L1

Some concepts to clarify before discussing the new techniques themselves.

What is a shrinkage estimator? (Ch. 7 intro)

What does regularization refer to? (7.3)

(ex. regression)

Describe the use of a training/test set in analysis.

The new techniques will involve tuning parameters L^2 that must be chosen, often using cross-validation. (CV)
What is your understanding of CV? For example, describe what happens in a 10-fold CV. (Ch 16, 16.1)

The jackknife approach, also called Leave-One-out-CV (LOOCV) is a special case of CV.
What makes it special?

Next is highlights about new techniques will implement in lab.

Ridge Regression

Matrix notation.

Review OLS

Regression setting. $y = X\beta + \varepsilon$ $\quad \varepsilon \sim \_\_(\_\_, \sigma^2 I)$? ?

Find $\hat{\beta}$ as $\hat{\beta} = (X'X)^{-1}X'Y$. The $\hat{\beta}$ is found to

________ the residual sum of squares, RSS.

$$RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\right)^2 \Rightarrow \|y - X\beta\|^2$$

Ridge is designed to improve on the OLS solution.

Ridge ________ β's towards 0, using a penalty term, with ______ parameter λ.

Ridge minimizes $RSS + \lambda \sum_{j=1}^{p} \beta_j^2$.

This term is a shrinkage ______.

matrix form:

$$\|y - X\beta\|^2 + \lambda\|\beta\|^2$$

Have to pick λ.

$\lambda = 0$ is OLS. As you $\uparrow \lambda$, what happens to the β estimates?

Can ridge set any β's to 0?

The ridge penalty is an $\ell_{\_}$ penalty.

Remember to use standardized variables.

What is the Bayesian rationale for Ridge regression?

L4

How do we pick λ? Assume you have a train/test data split, and a set of possible λ's.

Hint: Cross - ________. Explain in a few sentences.

Ch 16 - Sparse Modeling and the LASSO

Key concepts - Regularization is required if $p > n$, b/c typical OLS will fail.

Techniques mentioned in chapter - Best subsets, forward (stepwise) regression, LASSO, LARS, elastic net

---

In Stat 230, you learned about best subsets, forward selection (forward stepwise), backward elimination, and stepwise regression. Briefly, recap how each of these variable selection procedures work, conceptually.

LASSO - Least Absolute ______ and ______ Operator

- Want to do better than forward selection @ picking variables
- Want to do better than ridge in terms of shrinkage.

Ridge added an ℓ_2-penalty, $\lambda \sum_{j=1}^{p} \beta_j^2$.

Lasso adds an ___-penalty, $\lambda \sum_{j=1}^{p} |\beta_j|$.

This means LASSO <u>can</u> set β's to 0.
(Again, use standardized variables).
Often view the entire collection of solutions, over different values of t, where $\lambda \sum |\beta_j| \le t$.
Look @ Figure 16.5 (the LASSO path ex.) and make sure you can explain what it shows in your own words. Explanation:

Fitting LASSO Models L7

There is a lot of neat math explained here. I want to make sure a few points are clear.

What is the "active set" of variables for a fixed λ?

If the active set of variables does not change (including the signs of their β's) between $\lambda_1 < \lambda_2$, then $\hat{\beta}(\lambda)$ is ________ for $\lambda \in [\lambda_1, \lambda_2]$.

This means the coefficient profiles are continuous and ________ linear over the range of λ.

The "knots" occur when the active set (or a sign of a β) changes.

LAR - Least ________ Regression is an algorithm which capitalizes on the linearity properties above to fit the entire lasso regularization path.

The LAR algorithm needs an adjustment L8
to get the LASSO path to deal with
coefficient sign changes.

16.5 goes over some topics related to
computing LASSO solutions in other settings.
Some topics here could be future paper topics,
as they are not covered in our courses.

16.6 touches on some inference ideas.

16.7 has connections to techniques we will see
as we learn the other techniques

Last technique - Elastic net (in 16.5)

Elastic net bridges the gap between
________ and the LASSO b/c it uses a
________ term of the form:

$$P_\alpha(\beta) = \frac{1}{2}(1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1$$

Your text uses the <u>glmnet</u> package for
fitting and we will see it in lab.