How do you compute the PLS directions?

1) The first direction Z_1 is computed by setting each ϕ_{j1} equal to β_j from a simple linear regression of Y onto X_j, so

$$Z_1 = \sum_{j=1}^{p} \phi_{j1} X_j$$

is ~~so~~ strongly correlated with the response.

Okay, I don't understand this

2/21/23 Day 15

I'm gonna skip partial least squares for the time being.

6.4: Considerations in High Dimensions

6.4.1: High dimensional data

In the low dimensional setting, $n \gg p$. But in modern problems $p \gg n$, the high dimensional setting

- The bias-variance tradeoff and overfitting become especially important in this setting.

6.4.2: What Goes Wrong in High Dimensions

Here, we discuss least squares regression but this also applies to logistic regression, LDA, etc.

- Least squares produces a perfect fit when $p \geq n$, regardless of if there is really a linear relationship or not.
- A perfect fit almost certainly means overfitting to the data
- If $p > n$ or $p \approx n$, then linear regression is too flexible
- Including more predictors increases the variance of the coefficient estimates

6.4.3: Regression in High Dimensions

The methods of Chapter 6 are useful in the high dimensional setting

---

Can I use the lasso w/ Cox proportional hazards?

1) Regularization is important in high dimensional problems
2) Appropriate tuning parameter selection is crucial.
3) The test error tends to increase as the number of predictors increases, unless the additional features are truly associated with the response.

6.4.4: Interpreting Results in Higher Dimensions

- Multicollinearity is a huge problem in this setting — so you can end up with models w/ different predictors, (though it could be that both models produce accurate predictions).

- Be careful reporting errors since overfitting is so likely. NEVER use SSE, p-values, R^2, or measures of fit to the training data
- Use independent test set or cross validation errors instead.

Ch. 7: Moving Beyond Linearity

- Linear models are easily interpretable, but limited in predictive power.
- In this chapter, we discuss simple model extensions such as
  - Polynomial regression
  - Step functions
  - Regression splines
  - Smoothing splines
  - Local regression
  - Generalized additive models

  (Polynomial regression through Local regression: single predictors; Generalized additive models: multiple predictors)

7.1: Polynomial Regression

$$y_i = \beta_0 + \beta_1 x_i + \cdots + \beta_d x_i^d + \varepsilon_i$$

In practice, $d \le 4$ because otherwise, it's too flexible, especially at the boundary of the X variable

What is the variance of a fit i.e. $Var \hat{f}(x_0)$. We can use the reported variance/covariance matrix of the $\hat{\beta}_j$ to figure this out. If $\hat{C}$ is the 5x5 covariance matrix and $l_0^T = (1, x_0, x_0^2, x_0^3, x_0^4)$, then

$$Var[\hat{f}(x_0)] = l_0^T \hat{C} l_0$$

Can easily get 2*SE curves from this.

7.2: Step Functions

Polynomial functions impose a global structure on the non-linear function of X. We can use step function to avoid imposing such a bias.

- Break X into bins
- Fit a different constant to each bin

This converts a continuous variable into an ordered categorical variable.

Cutpoints $c_1, \ldots, c_K$, and then variables

$$C_0(X) = I(X < c_1)$$
$$C_1(X) = I(c_1 \leq X < c_2)$$
$$\vdots$$
$$C_{K-1}(X) = I(c_{K-1} \leq X < c_K)$$
$$C_K(X) = I(c_K \leq X)$$

Then fit the model

$$y_i = \beta_0 + \sum_{j=1}^{K} \beta_j C_j(x_i) + \varepsilon_i$$

Unless the predictor has natural breakpoints, piecewise constant functions can miss the action.

7.3: Basis Functions

Polynomials & step functions are special examples of the basis function approach. Have

$b_1(X), \ldots, b_K(X)$

and fit the linear model

$$y_i = \beta_0 + \sum_{j=1}^{K} \beta_j b_j(x_i) + \varepsilon_i$$

All the tools of chapter 3 are available in this setting.

Other basis functions include:

- Wavelets
- Fourier series
- Regression splines