# Stat 215A - Week 13

Zoe Vernon

Thanks to Rebecca Barter for sharing her slides

# Regularization: LASSO and Ridge Regression

# Regularization

Helps avoid overfitting the data.

A way to penalize as model complexity increases (e.g. many covariates in a linear regression).

Useful in modern data analysis when we have a lot of predictors compared to the number observations.

May introduce bias, but greatly reduces variance (overall reducing the MSE).

# Ridge Regression

Good when we have sparsity (e.g. we believe a lot of coefficients should be 0 or close to 0)

$L_2$ penalty that shrinks estimated coefficients in least squares towards 0 (although not all the way to 0).

$$\hat{\beta}^{\text{ridge}} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left\{ ||y - X\beta||_2^2 + \lambda ||\beta||_2^2 \right\}$$

$\lambda$ is a tuning parameter, when $\lambda$ is zero we return to OLS, as it increases we impose more penalty.

# LASSO

Good when we have sparsity (e.g. we believe a lot of coefficients should be 0 or close to 0)

$L_1$ penalty that causes some estimated coefficients to be exactly 0, which makes for easier interpretation than ridge regression.

$$\hat{\beta}^{\text{lasso}} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left\{ ||y - X\beta||_2^2 + \lambda ||\beta||_1 \right\}$$

Again, small $\lambda$ mean our estimates are closer to OLS and larger values force more coefficients to 0.

# Elastic net

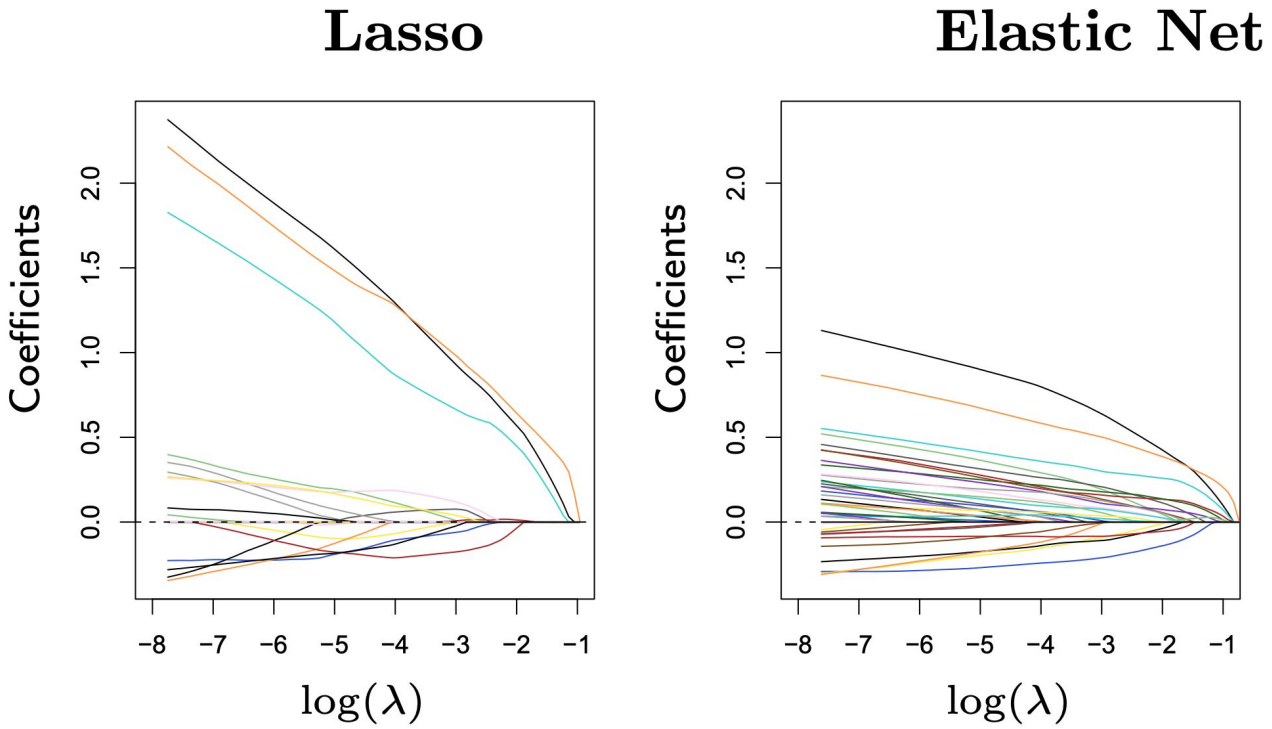
Combination of $L_1$ and $L_2$ penalties.

$$\hat{\beta}^{\text{elastic net}} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left\{ ||y - X\beta||_2^2 + \lambda_1 ||\beta||_1 + \lambda_2 ||\beta||_2^2 \right\}$$

Good when there is a lot of correlation between variables.

- LASSO will tend to select one of the correlated variables
- Ridge tends to shrink the coefficients for these correlated values towards each other.

Source: http://www2.stat.duke.edu/~banks/218-lectures.dir/dmlect9.pdf

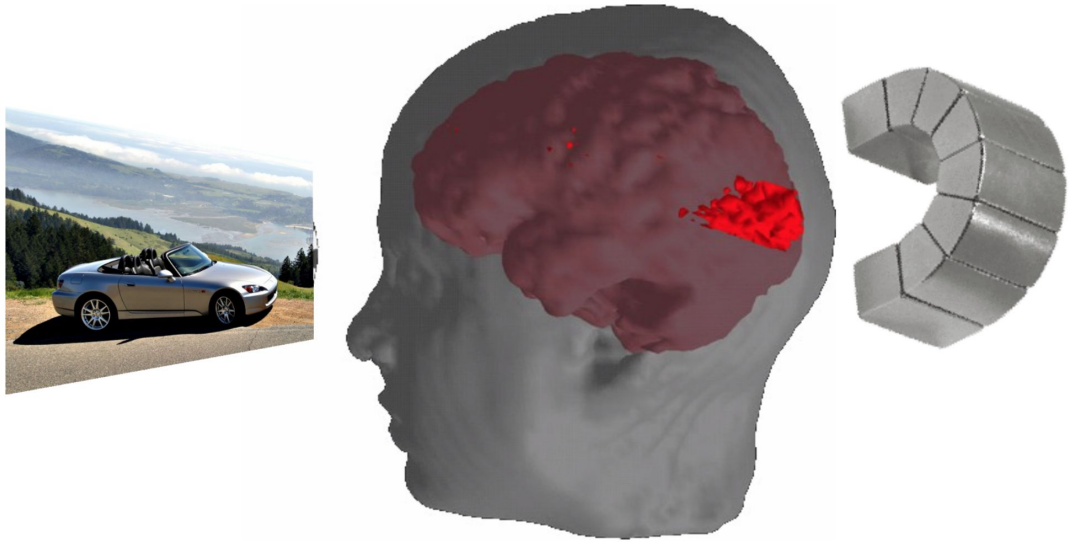


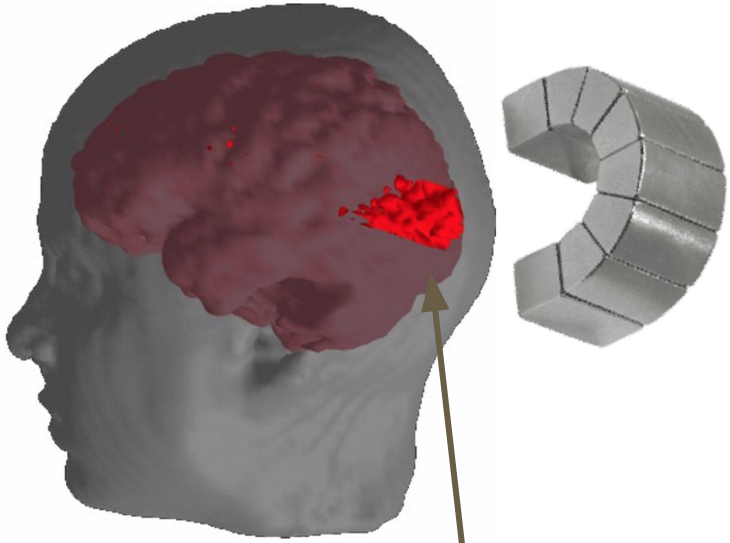

Source: http://www2.stat.duke.edu/~banks/218-lectures.dir/dmlect9.pdf

# Final Project (due Dec 7)

# Final project in a nutshell

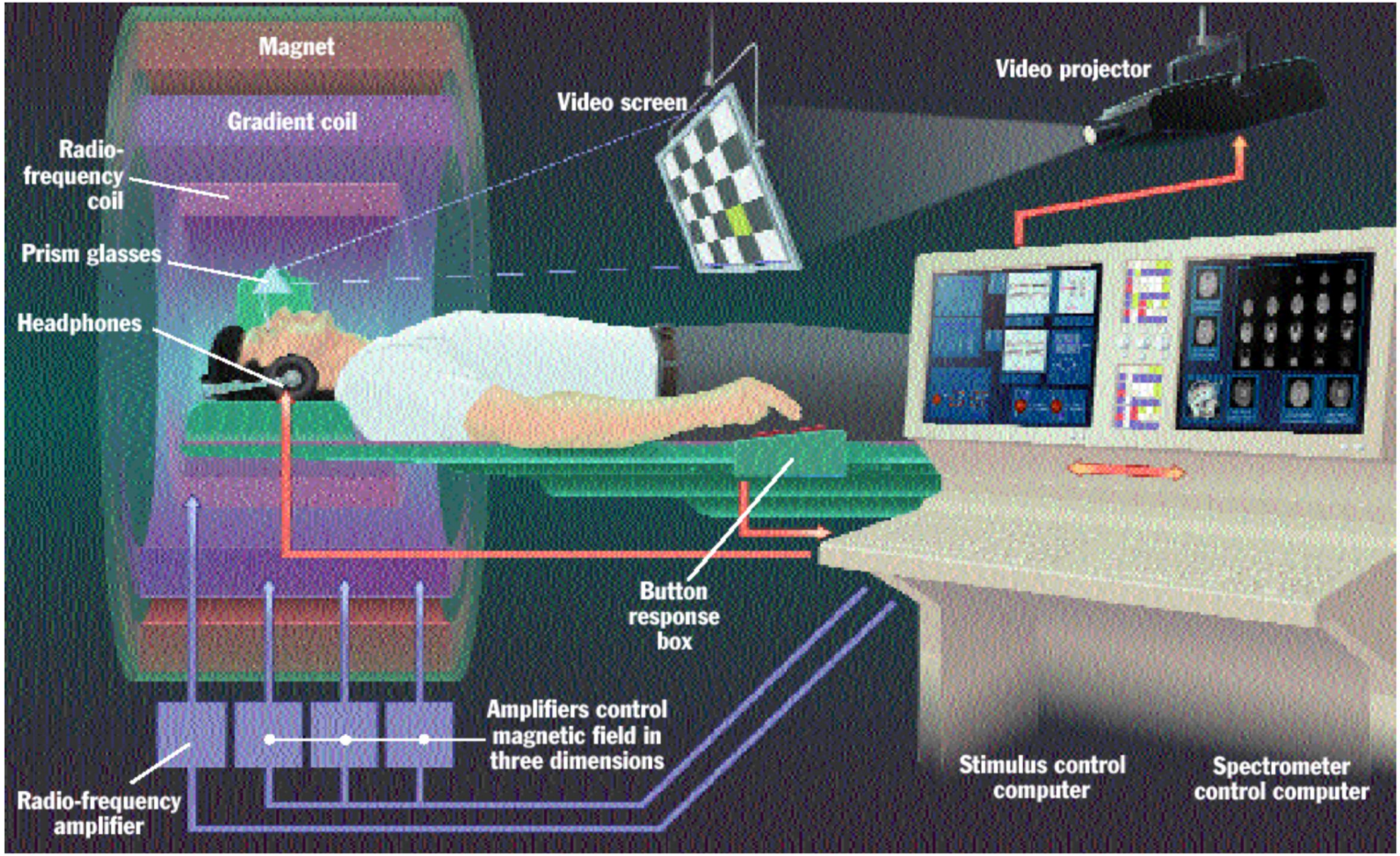
Predict the brain's response to images

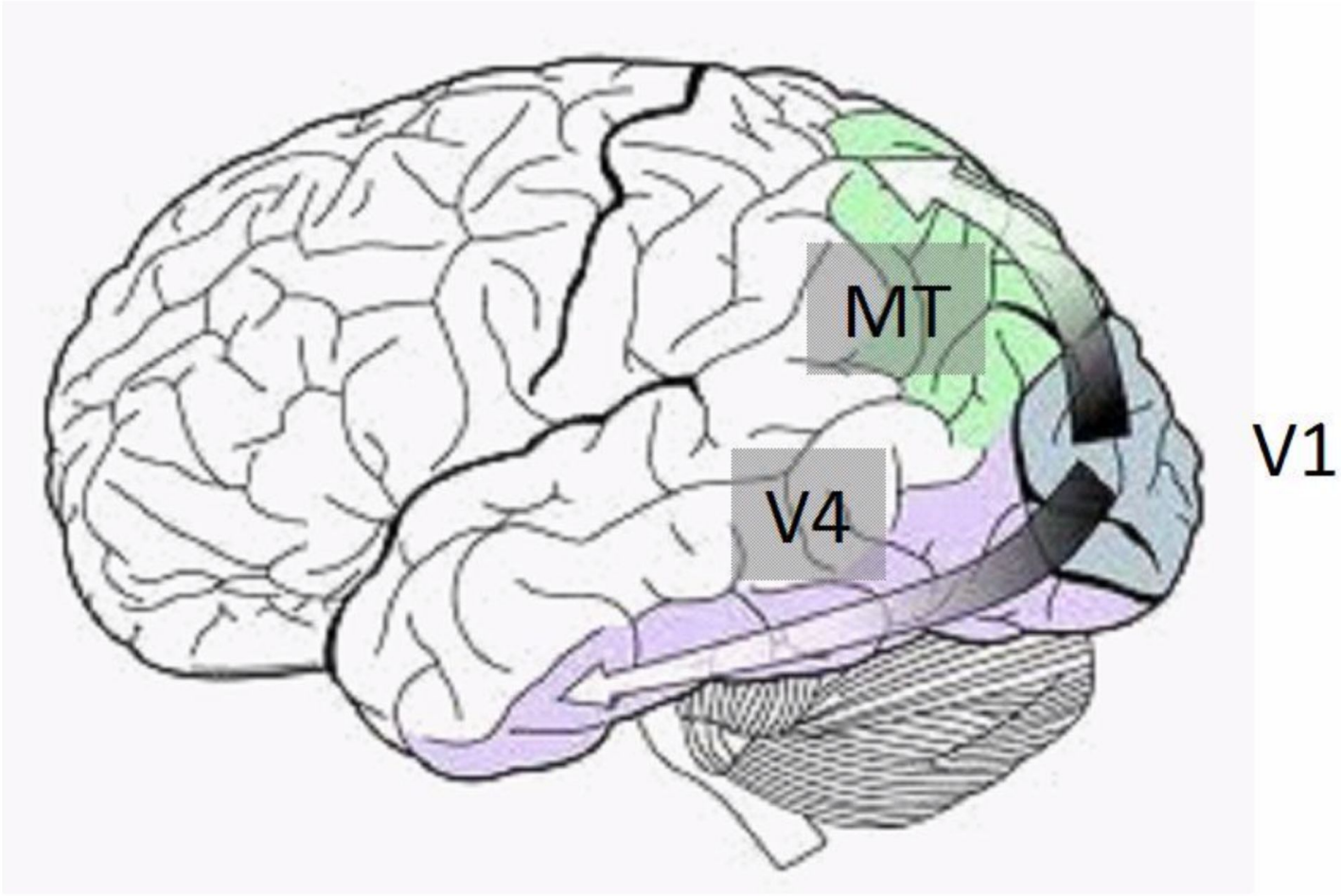




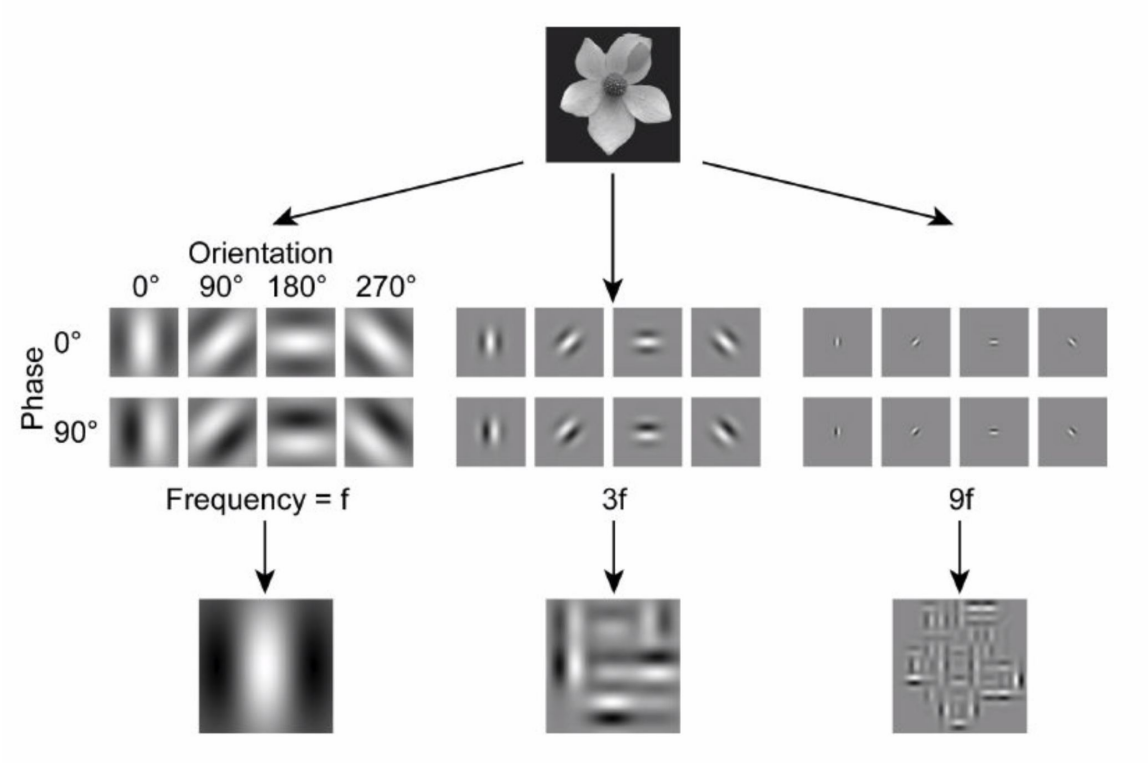
X = stimulus

Y = response (20 locations in V1 region)

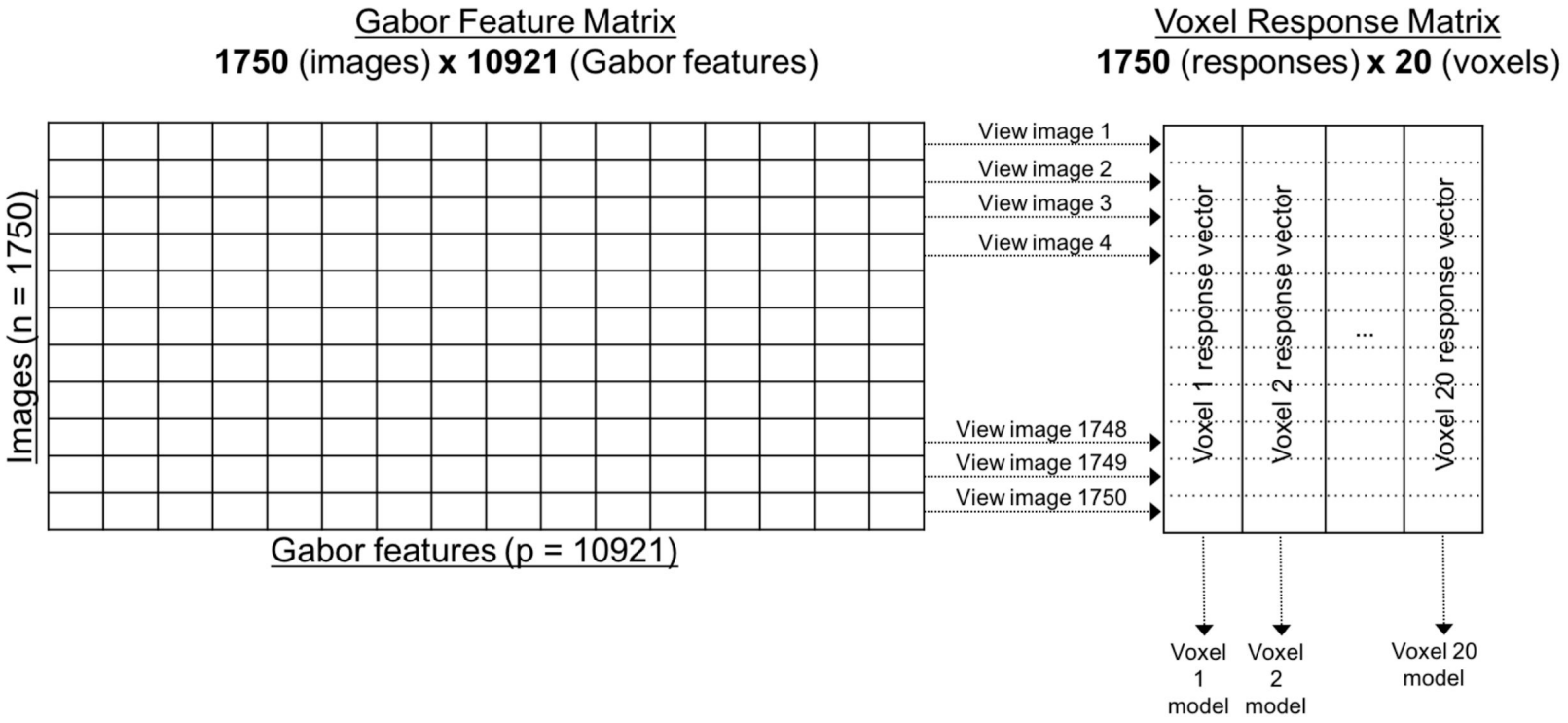

Magnet
Gradient coil
Radio-frequency coil
Prism glasses
Headphones
Video screen
Video projector
Button response box
Amplifiers control magnetic field in three dimensions
Radio-frequency amplifier
Stimulus control computer
Spectrometer control computer


MT
V4
V1

# Gabor transform



# The data


Gabor Feature Matrix
1750 (images) x 10921 (Gabor features)
Voxel Response Matrix
1750 (responses) x 20 (voxels)
Images (n = 1750)
Gabor features (p = 10921)
View image 1
View image 2
View image 3
View image 4
View image 1748
View image 1749
View image 1750
Voxel 1 response vector
Voxel 2 response vector
...
Voxel 20 response vector
Voxel 1 model
Voxel 2 model
Voxel 20 model

# Looking at the data

Example in fmri_example.R

# Caret tutorial

See caret_tutorial.Rmd