STATISTICAL AND ANALYSIS ISSUES

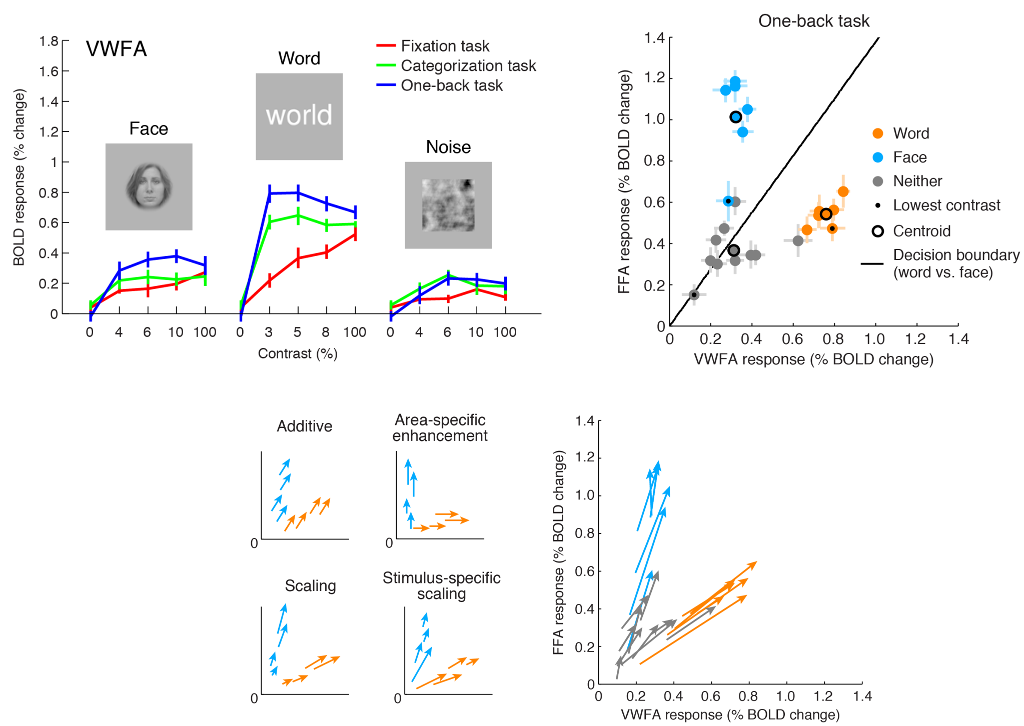
Overview

- Here we touch upon fundamental concepts pertaining to modern statistical and computational analysis
- These are general-purpose tools at our disposal that can be applied to different fMRI analysis approaches
- Many of these are nonparametric methods that avoid risky and cumbersome assumptions (at the expense of computational time)

Some resources:
Statistics materials: https://www.cmrr.umn.edu/~kendrick/statsmatlab/
Statistics blog: http://randomanalyses.blogspot.com

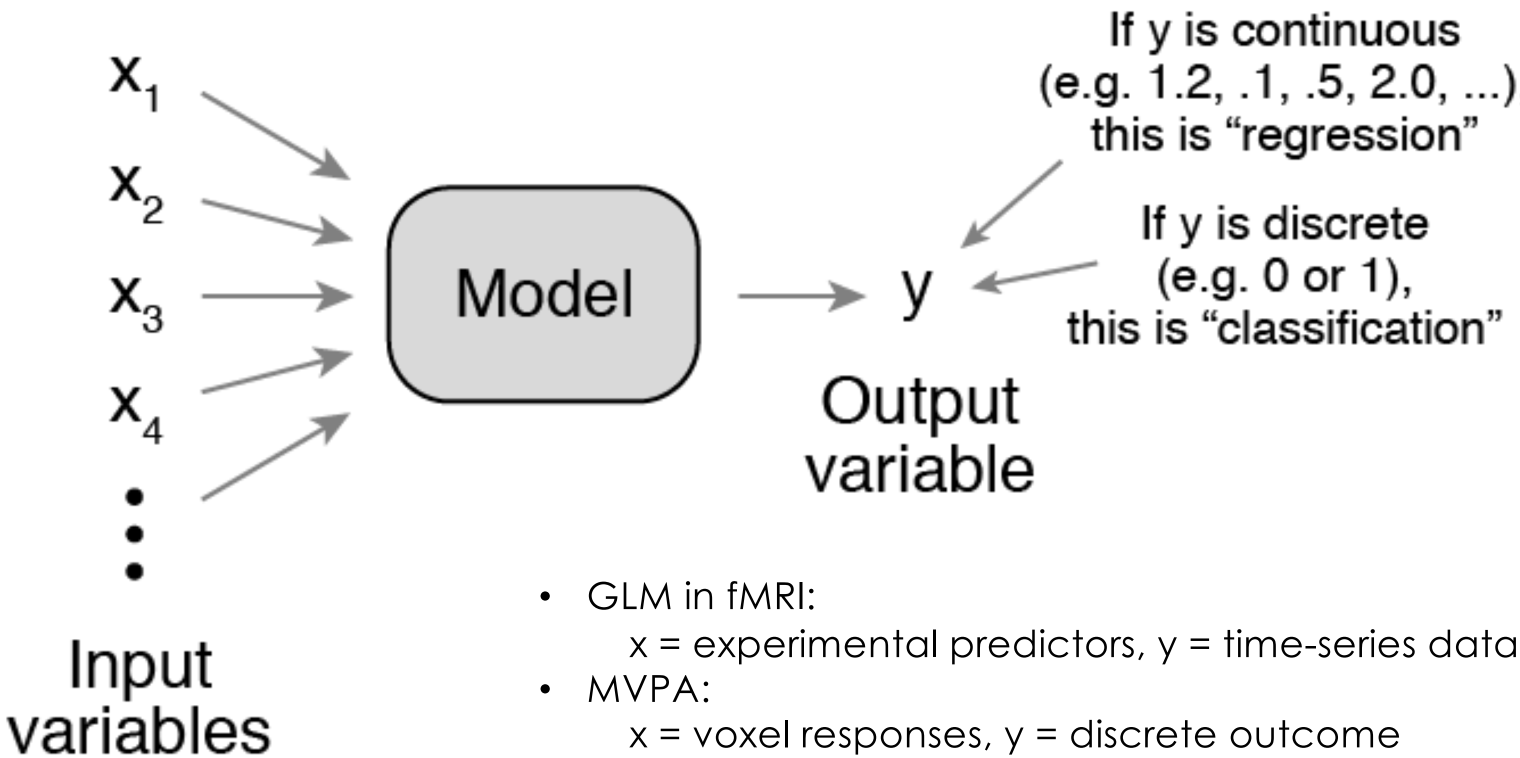
Before trying to analyze data, look at it!

- Good visualization can do wonders.
- And it is not necessarily easy. Take the time to do it.



from Kay & Yeatman, *eLife*, 2017

Supervised learning



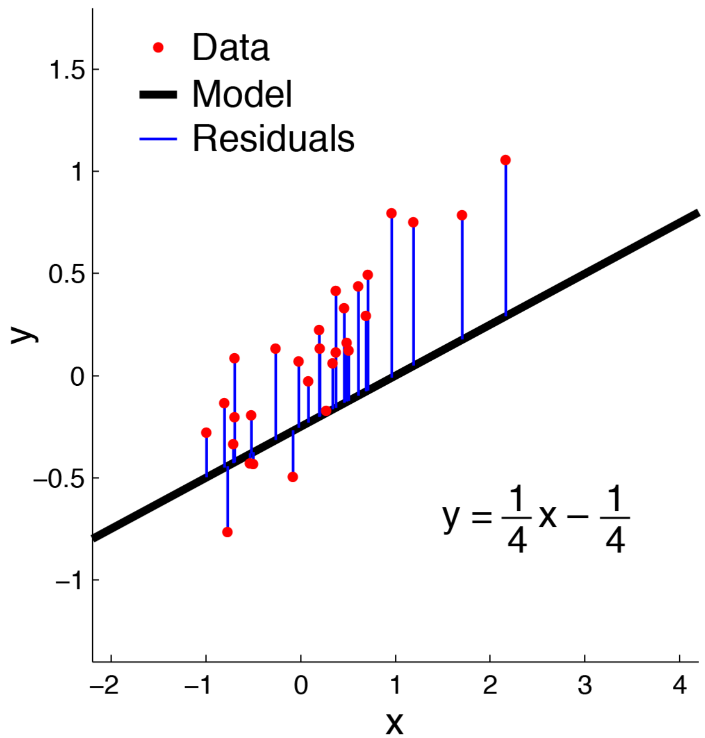
- GLM in fMRI:
  x = experimental predictors, y = time-series data
- MVPA:
  x = voxel responses, y = discrete outcome
- Encoding models:
  x = stimulus features, y = measured response

Regression

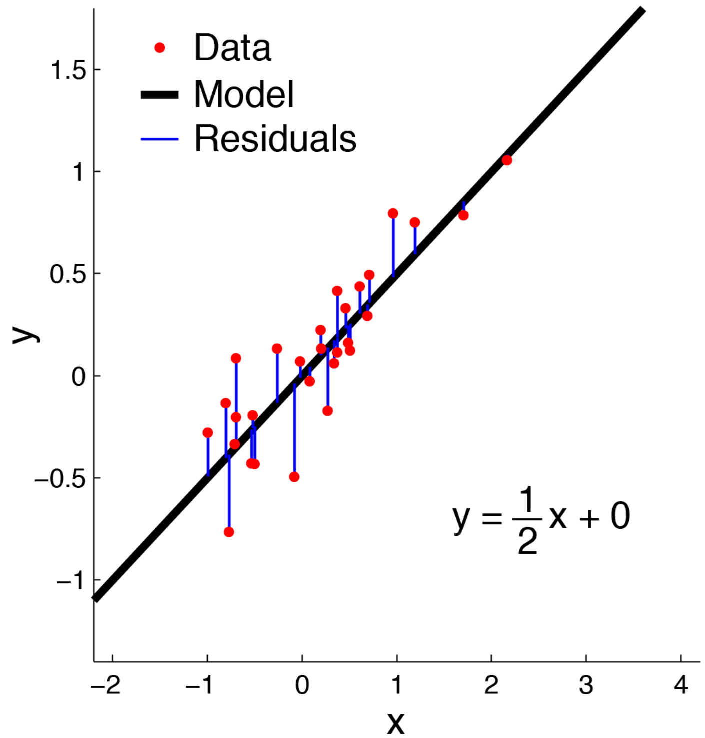
- Probably the most important topic to truly understand
- Use a weighted sum of regressors (input) to fit some data (output)

$$\text{squared error} = \sum_{i=1}^{n} (d_i - m_i)^2$$

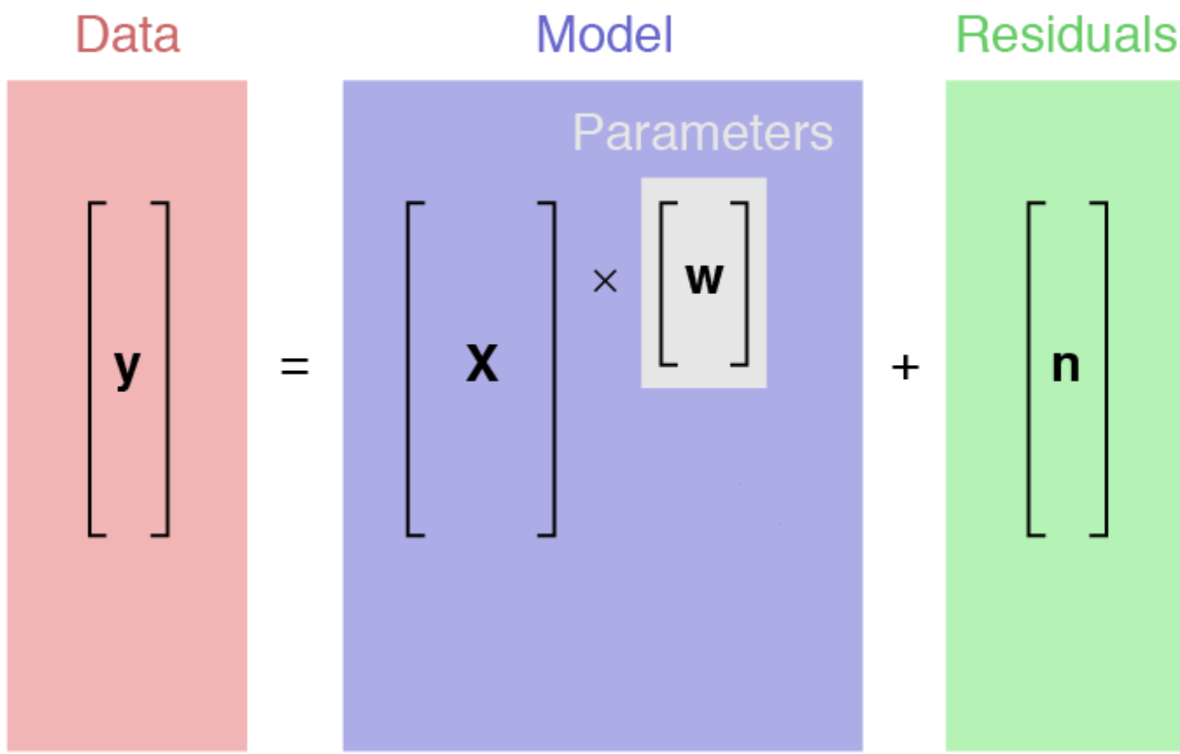
High squared error



Low squared error

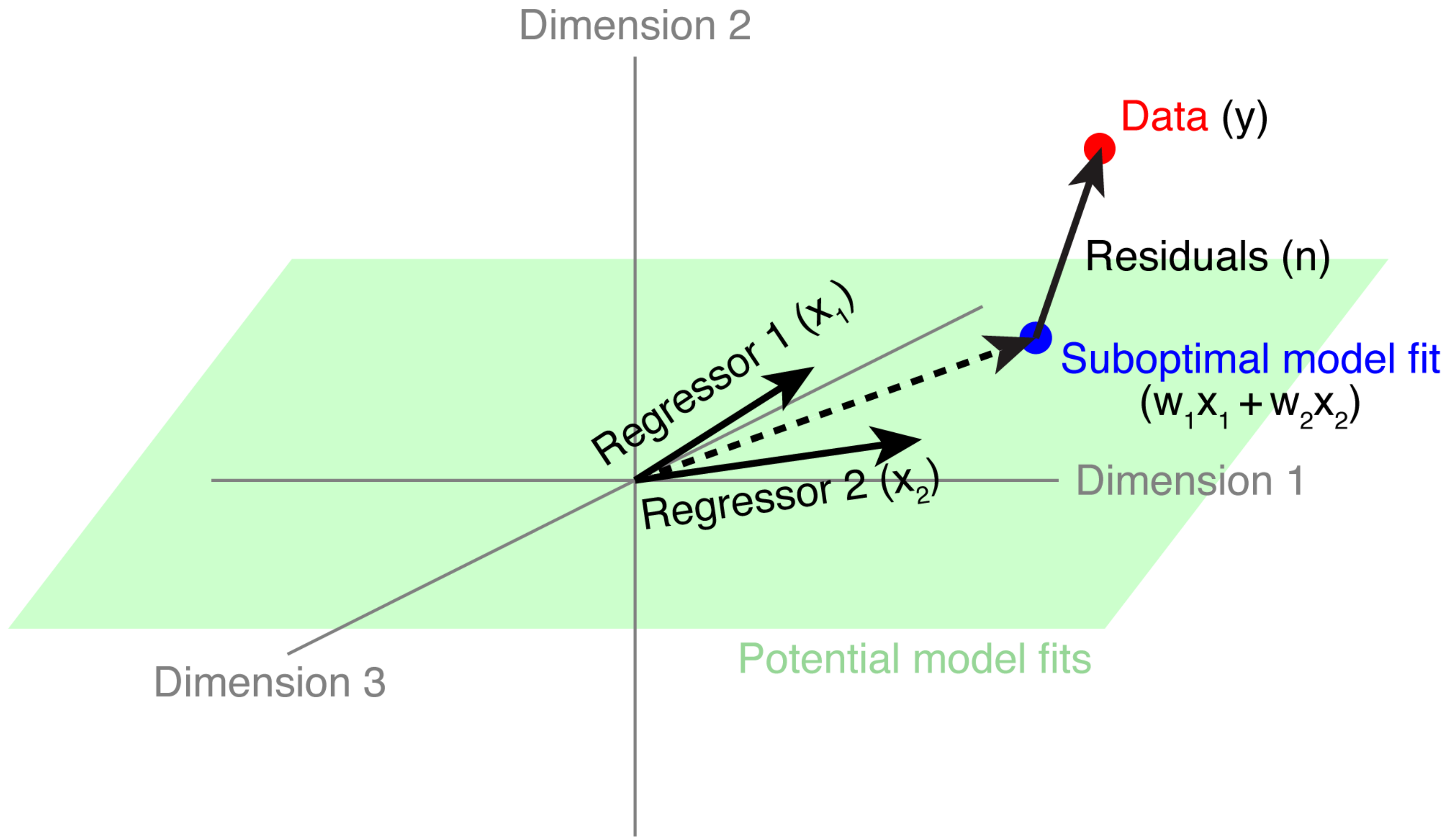


Matrix representation of linear models



$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{n}$$

Geometric interpretation of linear regression



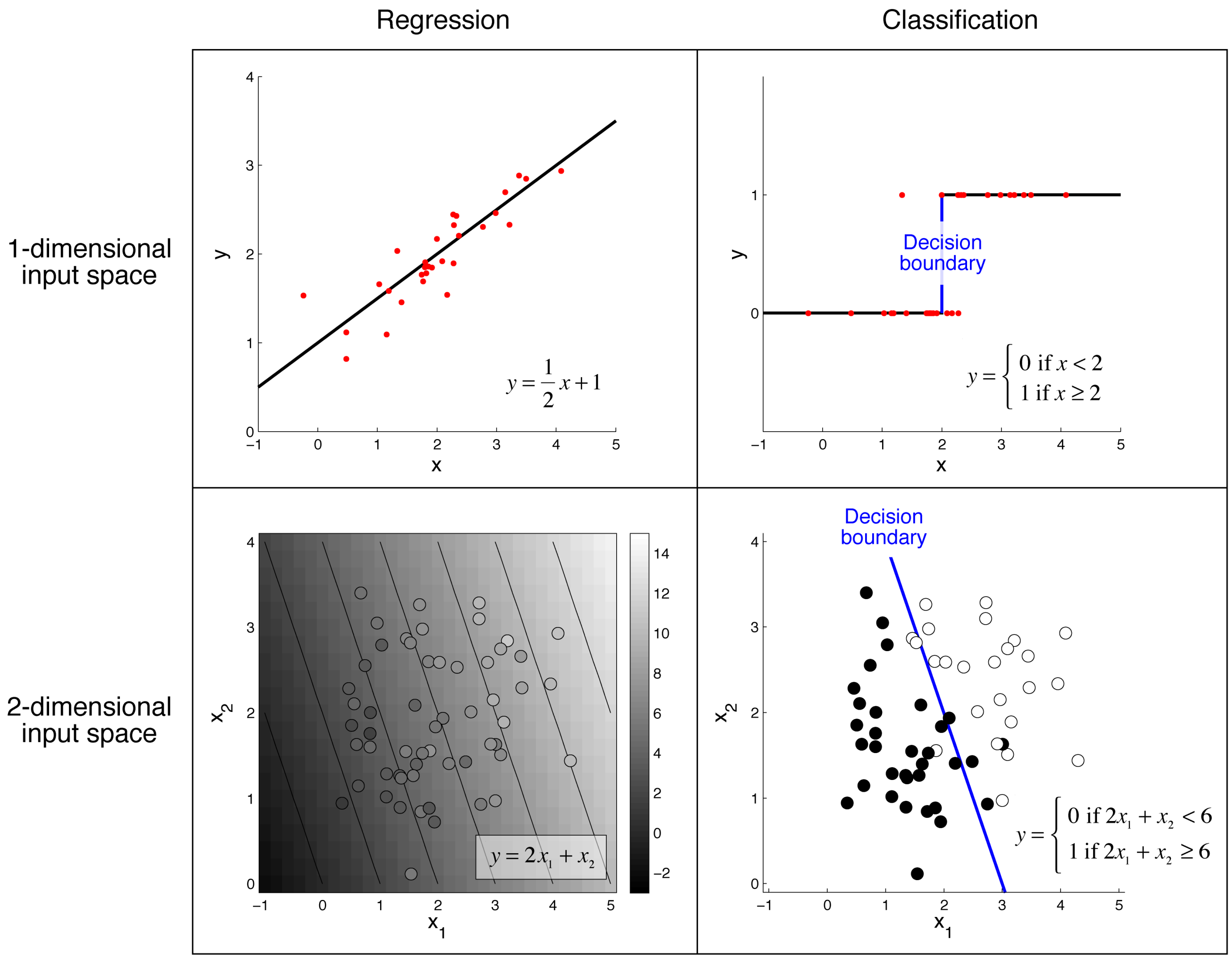
Classification

- In the linear case, use a weighted sum of predictors (input) followed by a threshold to fit the data (output)
- Nonlinear classifiers are similar, except that we are no longer restricted to weighted sums of the predictors

Linear classification model

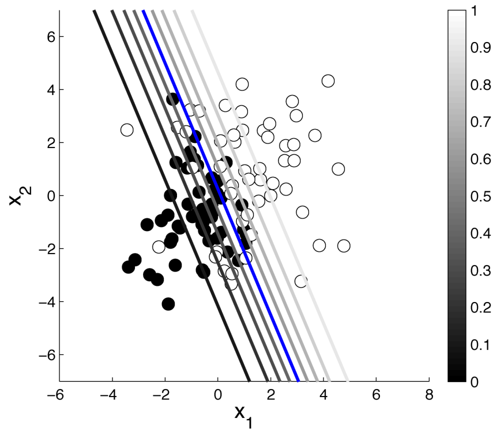
$$y = \begin{cases} 0 \text{ if } \sum_{i=1}^{n} w_i x_i < c \\ 1 \text{ if } \sum_{i=1}^{n} w_i x_i \geq c \end{cases}$$

Comparing regression and classification

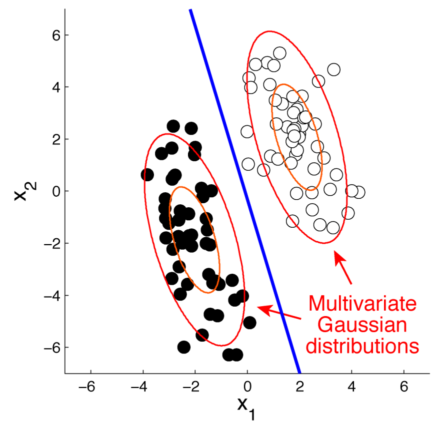


Some classification techniques

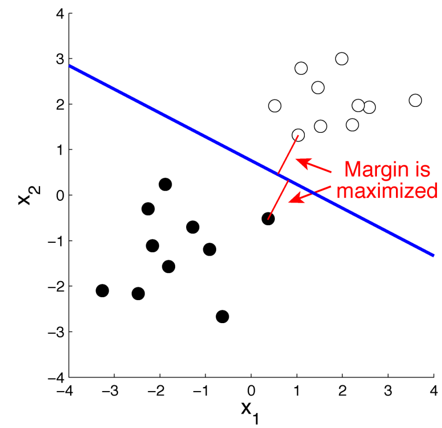
Logistic regression



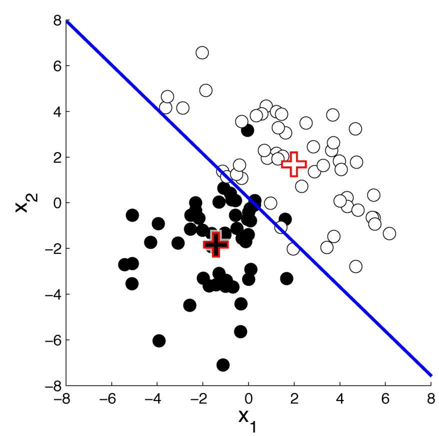
Linear discriminant analysis (LDA)



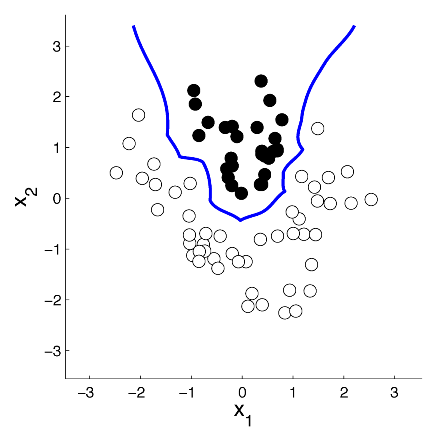
Support vector machines (SVM)



Nearest-prototype classification



Nearest-neighbor classification



Both regression and classification involve model building

Model specification
(what type of model to use?)

Model fitting
(how do we estimate the model parameters?)

Model accuracy
(how well does the model describe the data?) ← Cross-validation!

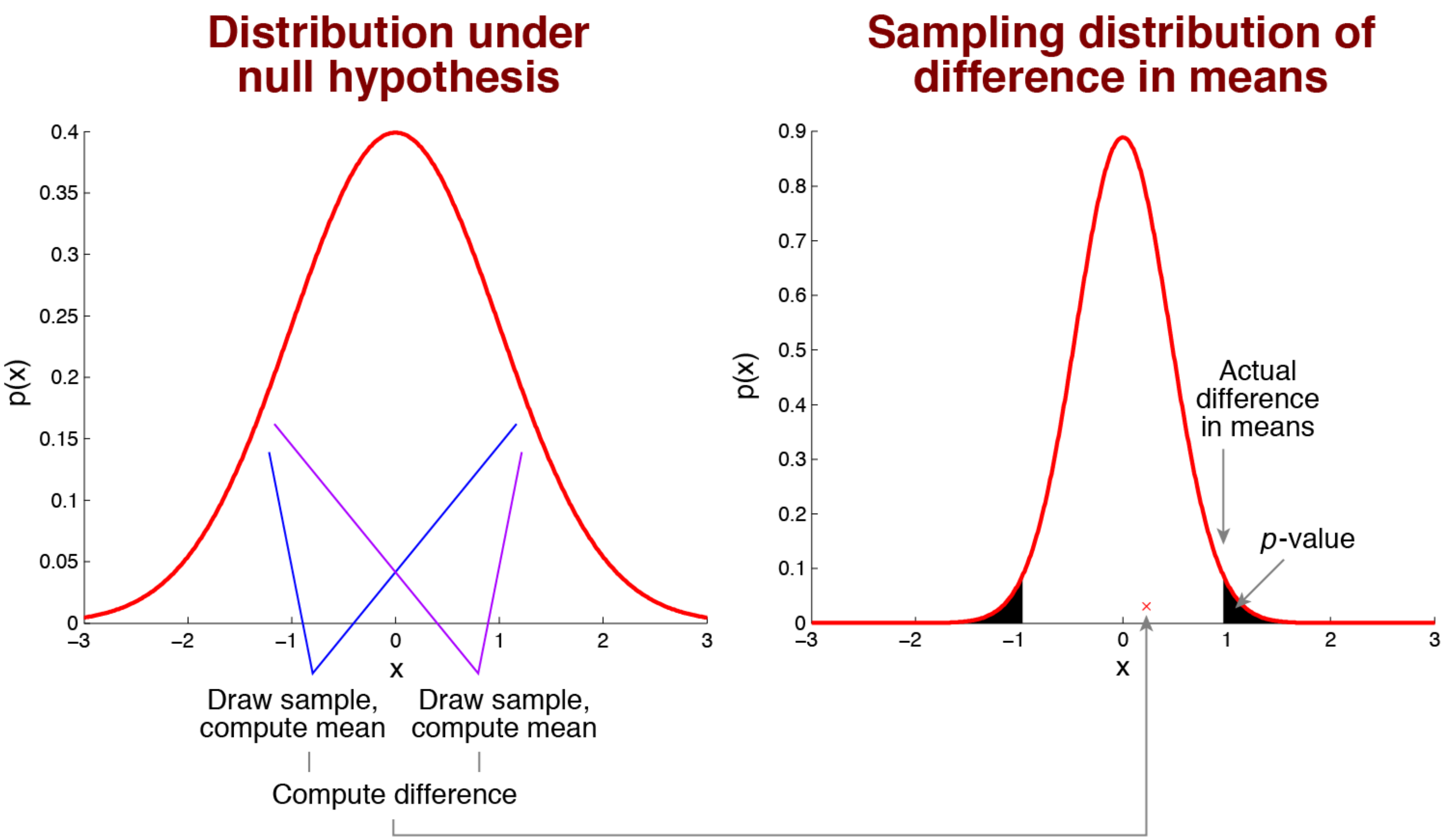
Model reliability
(how stable are the parameter estimates?) ← Bootstrapping!

Resampling methods

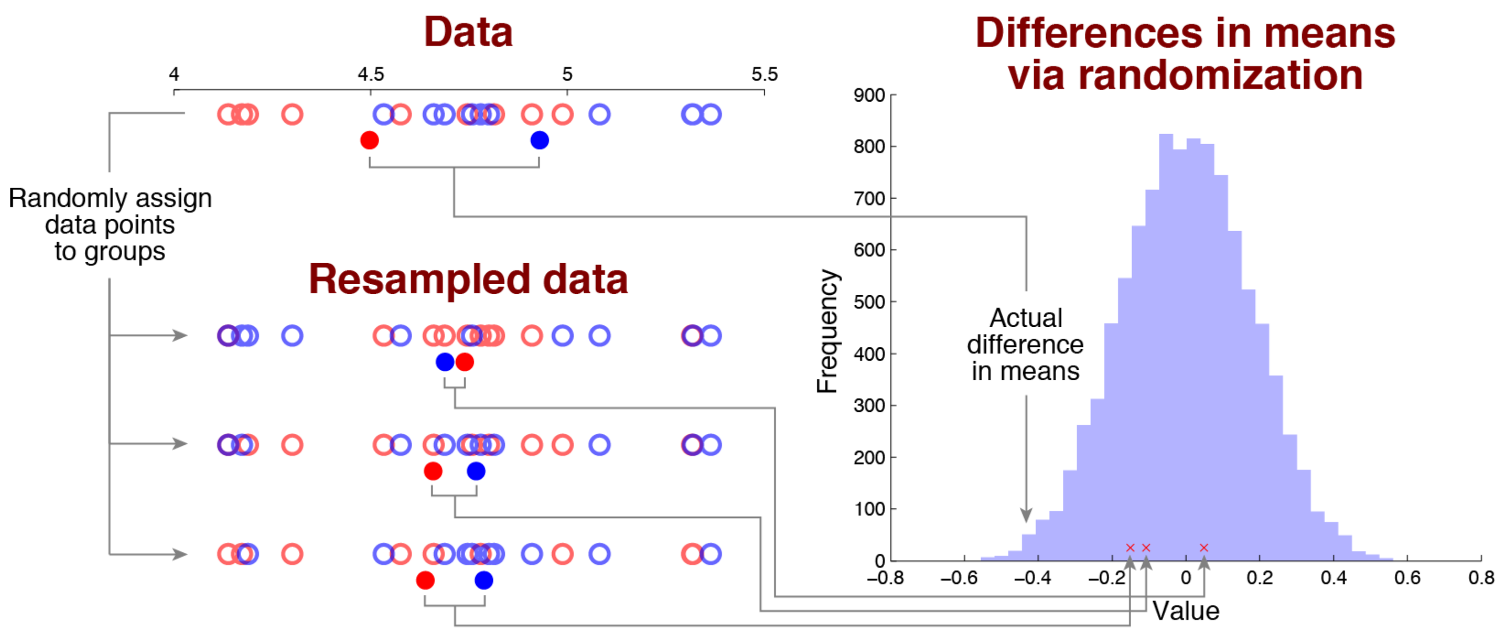
- With computational power, we can carry out powerful resampling methods.
- Examples to be covered:
  - Randomization/permutation
  - Bootstrapping
  - Cross-validation

Randomization/permutation

- Useful for NHST
- Idea: Randomize or permute the data according to a null hypothesis. Then calculate a *p*-value based on how unlikely the actual results are.
- Appealing features:
  - Easy
  - Nonparametric
  - Randomized data are well matched to the actual data
- Example: permute stimulus labels in order to determine which voxels exhibit a statistically significant level of selectivity
- Example: in a correlation of two variables, permute one variable to determine the significance level of the correlation



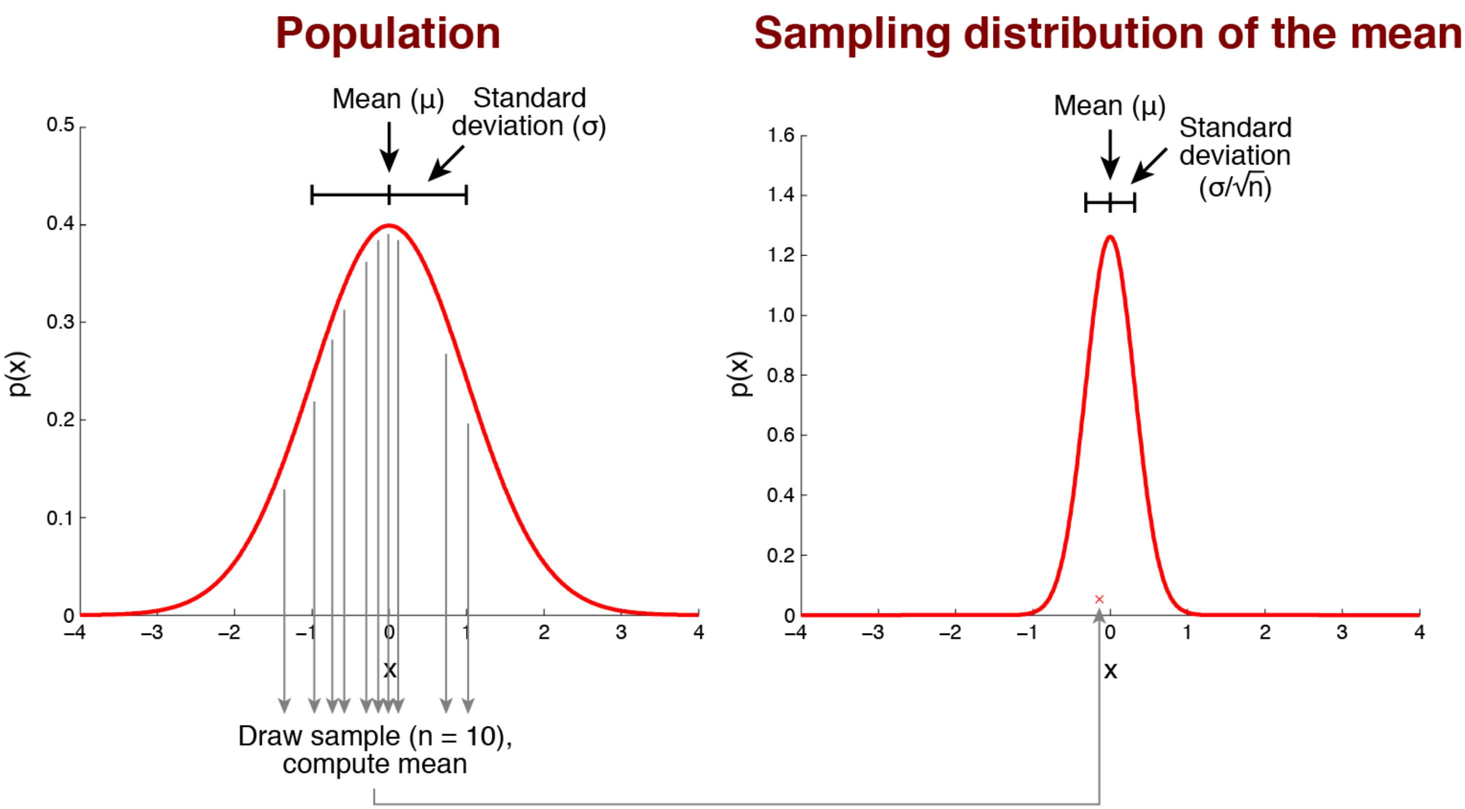
Distribution under null hypothesis
Sampling distribution of difference in means
p(x)
x
Draw sample, compute mean
Draw sample, compute mean
Compute difference
Actual difference in means
p-value

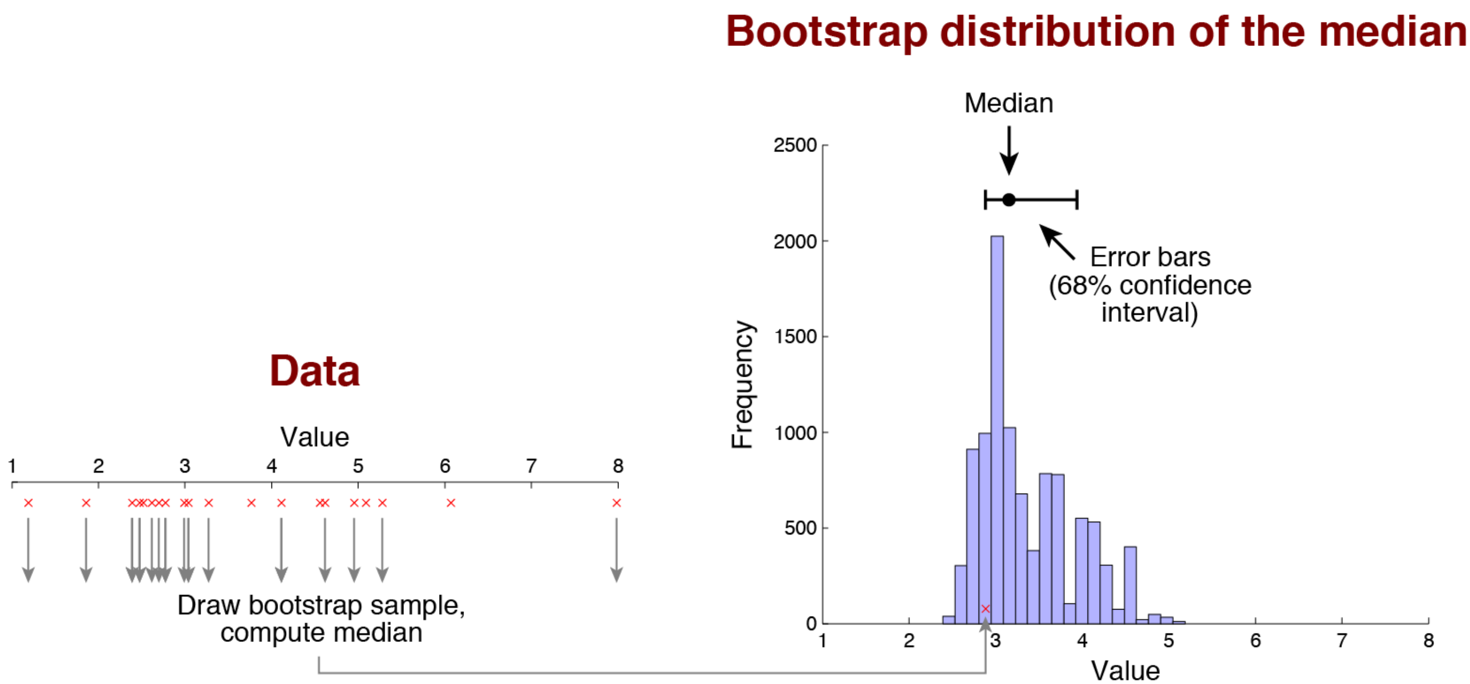


Data
4
4.5
5
5.5
Randomly assign data points to groups
Resampled data
Differences in means via randomization
900
800
700
600
500
400
300
200
100
0
Frequency
Actual difference in means
–0.8
–0.6
–0.4
–0.2
0
0.2
0.4
0.6
0.8
Value

Bootstrapping

- Idea:
  - Use the sampled (observed) data as a proxy for the population
  - Draw data points with replacement and perform the analysis for each bootstrap sample
- Provides an estimate of the reliability of analysis results
- Appealing features:
  - Easy
  - Nonparametric
  - Compatible with any analysis procedure
- Example: bootstrap subjects, or fMRI runs, or trials
- Note: Data points must be statistically independent. Cannot bootstrap voxels!! Cannot bootstrap volumes!!


Population
Sampling distribution of the mean
Mean (μ)
Standard deviation (σ)
Standard deviation (σ/√n)
p(x)
x
Draw sample (n = 10), compute mean

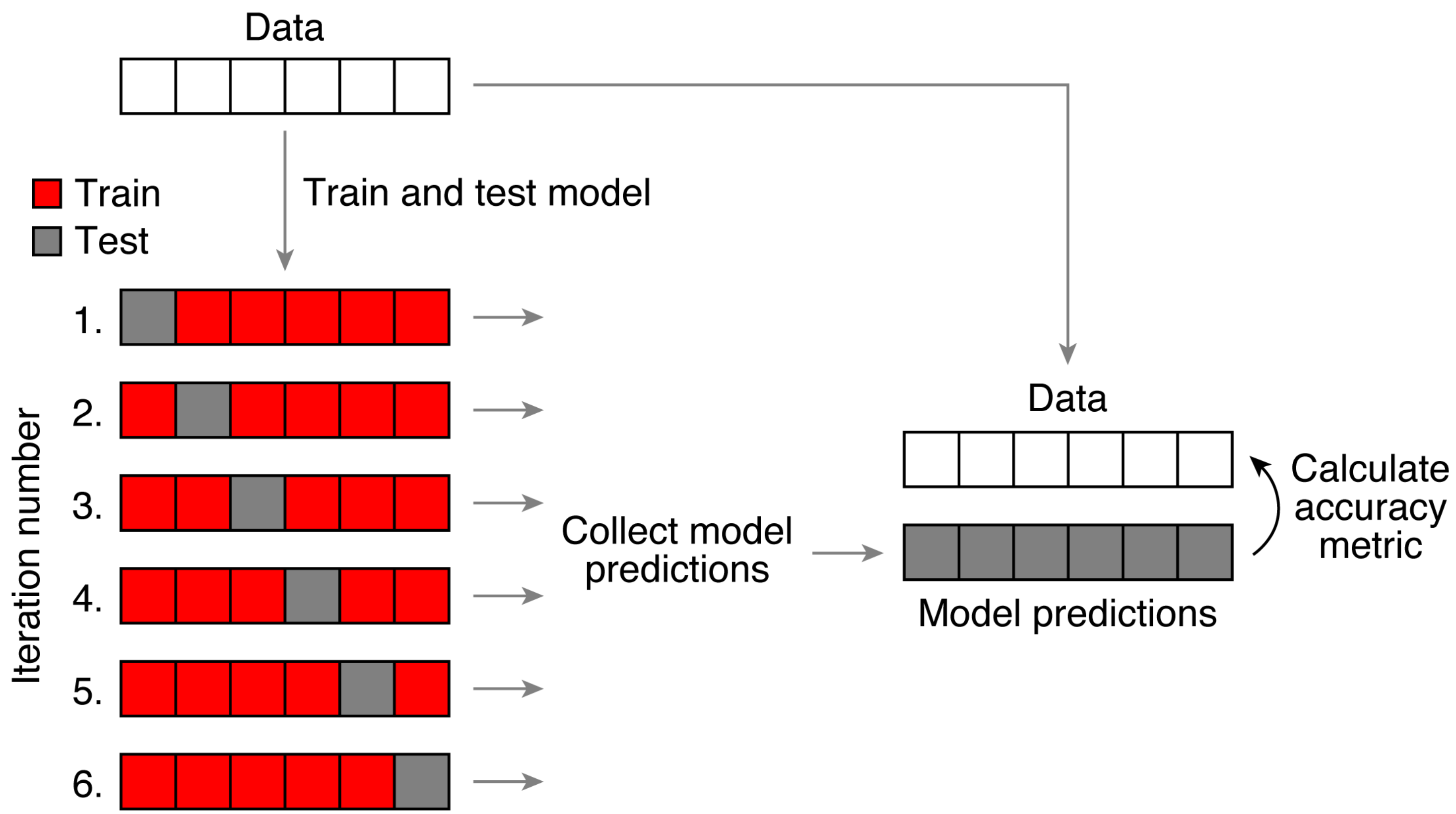


Bootstrap distribution of the median
Median
Error bars
(68% confidence
interval)
Frequency
2500
2000
1500
1000
500
0
1
2
3
4
5
6
7
8
Value
Data
Value
1
2
3
4
5
6
7
8
Draw bootstrap sample,
compute median

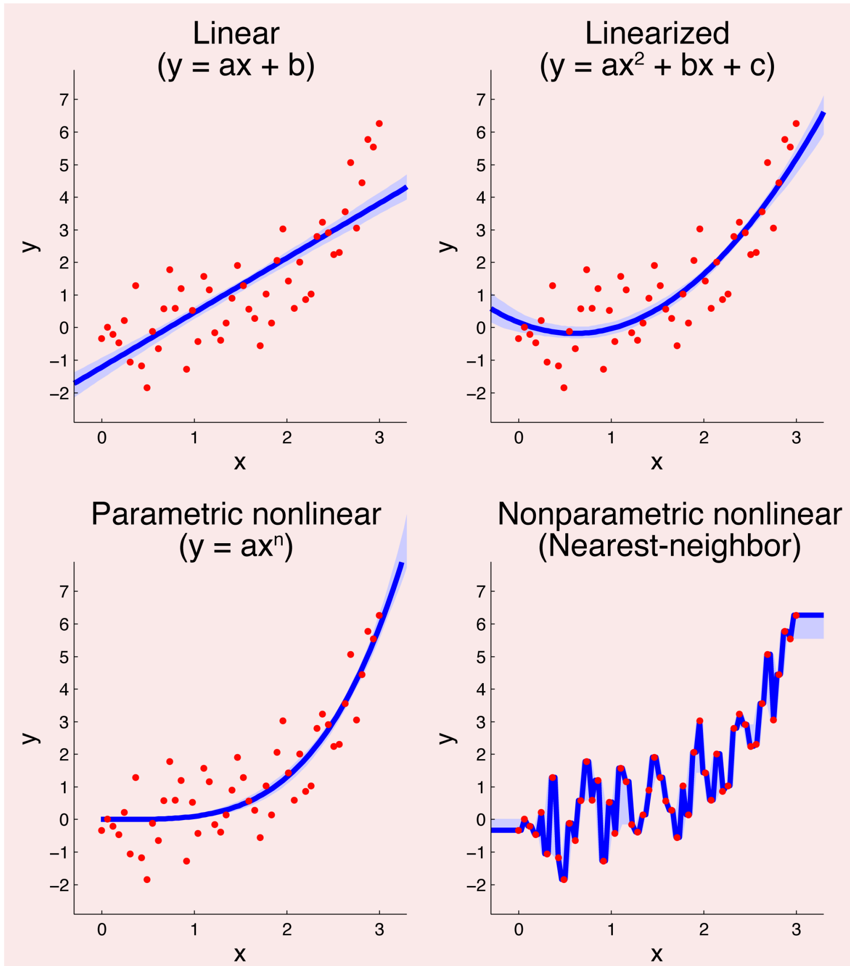
Cross-validation

- Idea:
  - Performance (accuracy) on training data is overly optimistic (biased)
  - Use left-out data to obtain an unbiased estimate of performance
- Basic procedure:
  1. A subset of data is left out ("testing data"),
  2. The remaining data are used to fit a model ("training data")
  3. The fitted model's performance is assessed on the left-out data.
- Some jargon:
  - estimation vs. validation
  - fit vs. predict
  - training data vs. testing data
- Many different cross-validation schemes (80/20 is common rule of thumb)
- Example: Cross-validate MVPA on left-out runs
- Example: Construct an atlas from N-1 subjects and test on subject N

Leave-one-out cross-validation

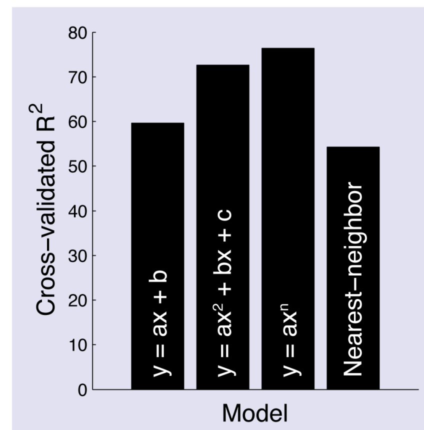


Bootstrapping and cross-validation applies to all models



Use bootstrapping to estimate model reliability

Use cross-validation to estimate model accuracy

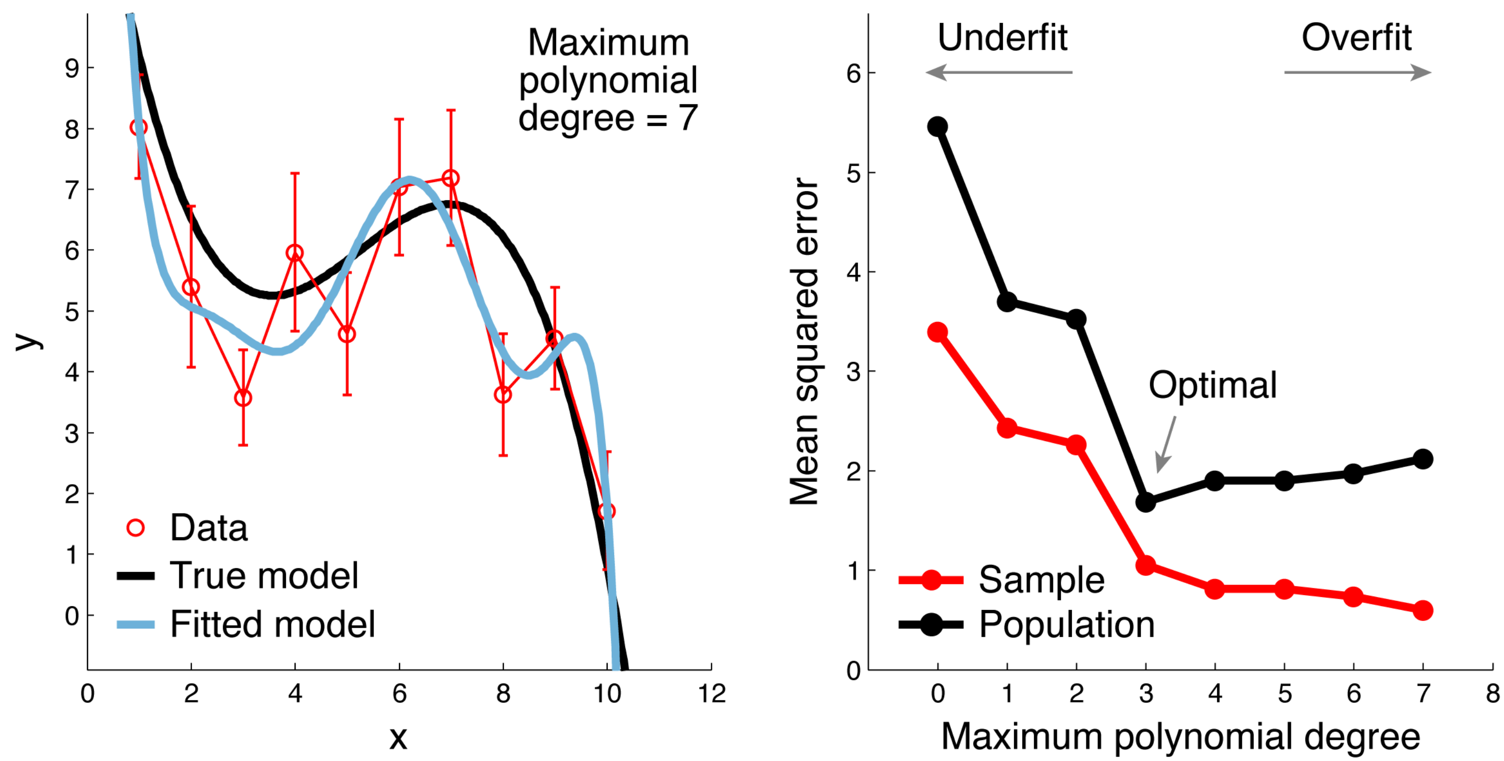


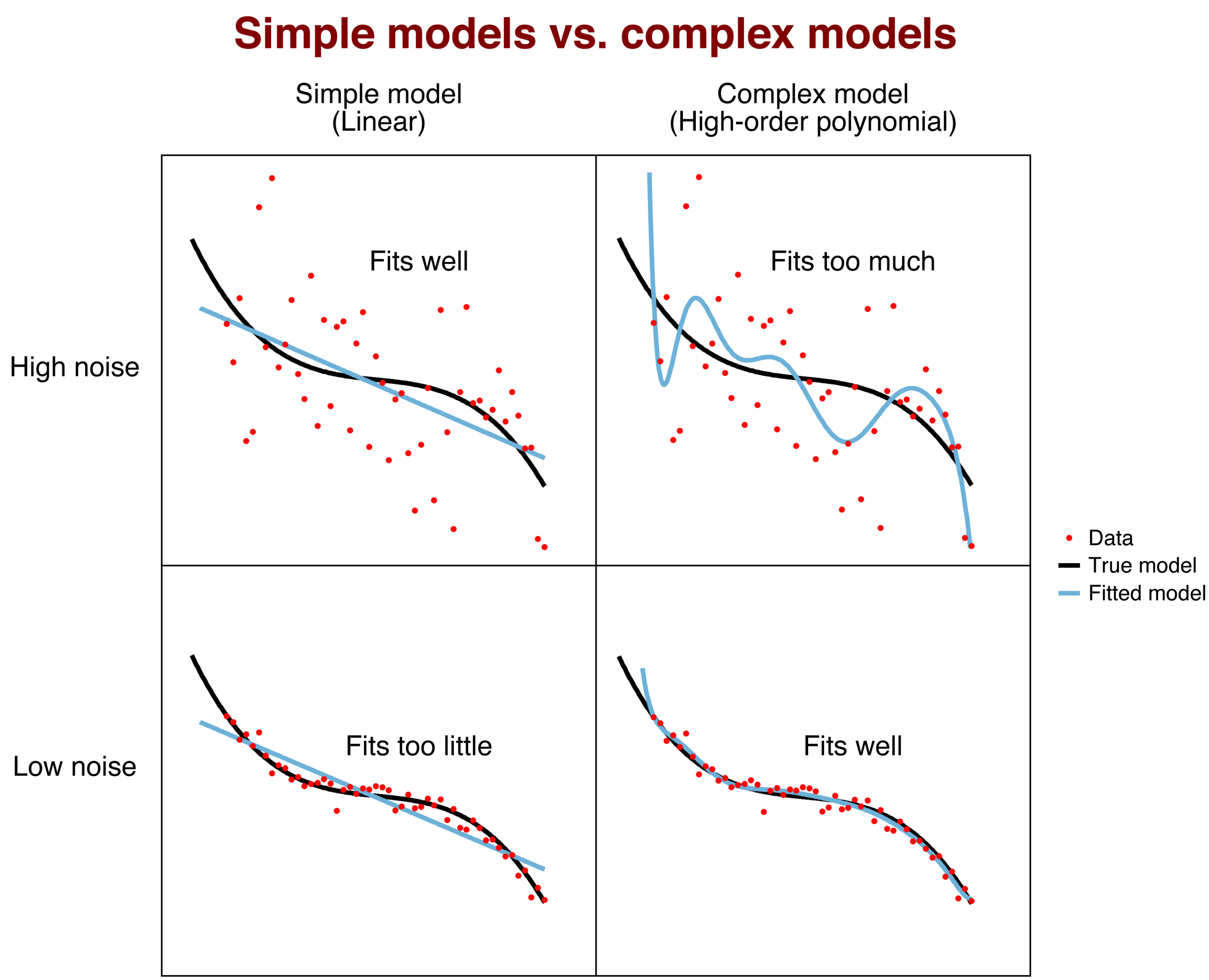
Overfitting

- Refers to a scenario where a model with lots of flexibility is allowed to fit large amounts of variability in the data.
- Such a model risks fitting the noise in the data and therefore may fail to generalize well to new data.
- Example: when the number of (non-collinear) regressors exceeds the number of data points, 100% of the variance can be fit, but will it really predict 100% of future data?

Overfitting

$$y = ax^7 + bx^6 + cx^5 + dx^4 + ex^3 + fx^2 + gx + h$$




Simple models vs. complex models
Simple model
(Linear)
Complex model
(High-order polynomial)
High noise
Low noise
Fits well
Fits too much
Fits too little
Fits well
Data
True model
Fitted model

Regularization

- Generally refers to a method that "smooths out" a solution or makes it otherwise more well-behaved
- A specific method of regularization is to incorporate additional constraints or penalties into a cost function
  - A classic example is ridge regression in which the sum of the squares of the weights is penalized
  - Another example is lasso in which the sum of the absolute values of the weights is penalized
- By imposing regularization, the **variance** in parameter estimates is reduced, at the cost of introducing **bias**
- Cross-validation can be used to determine the optimal level of regularization

General statistical and analysis issues that one should keep in mind

- Flexibility in pre-processing and analysis
  (do results depend on choice of parameters used?)
- Do the results depend on any arbitrary thresholds?
- Is there circularity in the analysis? (E.g., define a ROI based on a criterion, and then later "discover" that the ROI exhibits the behavior)
- Important to think about what entities are independent and what entities are not. Voxels are not independent.
- Could results be driven by artifacts correlated with the effect of interest (e.g., head motion, attention)?
- Can results be demonstrated for individual subjects, or only at the group level?
- How exactly are data combined across subjects (and how valid is the correspondence/registration)?