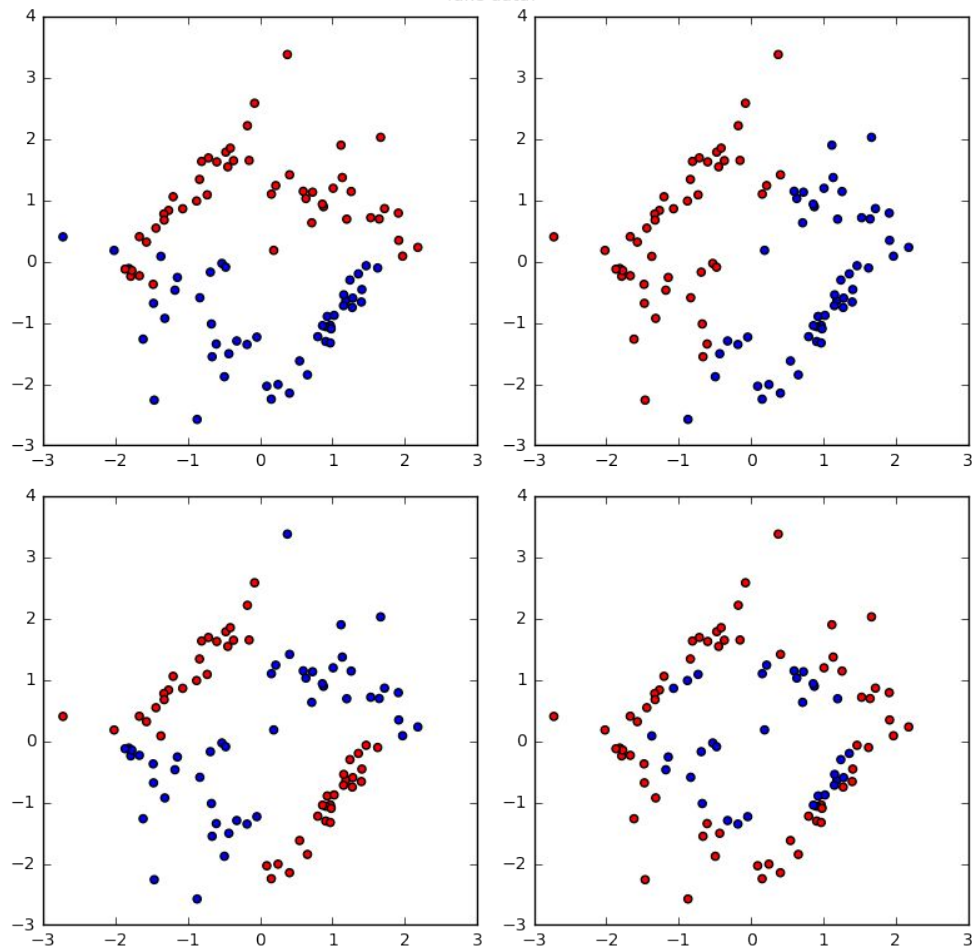# Model Comparison Guide

Pair discussion

- Briefly define & discuss the following terms
    - Supervised vs. Unsupervised learning
    - Regression vs. Classification
    - Parametric vs. Non-parametric model
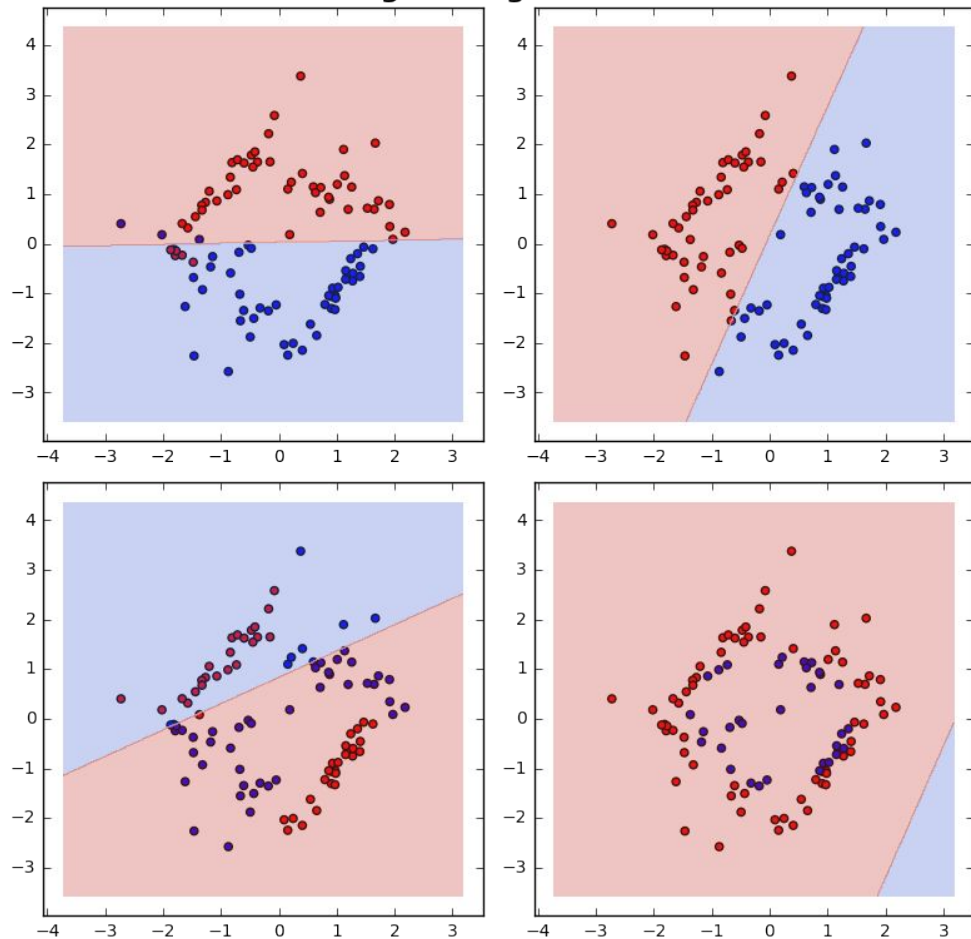   - Parameter vs. Hyperparameter

Individual exercise, then pair discussion

- What is a case where you would prioritize a model's *interpretability* over its *performance*? How about the converse?
- What kinds of models come to mind for these cases?

fake data!
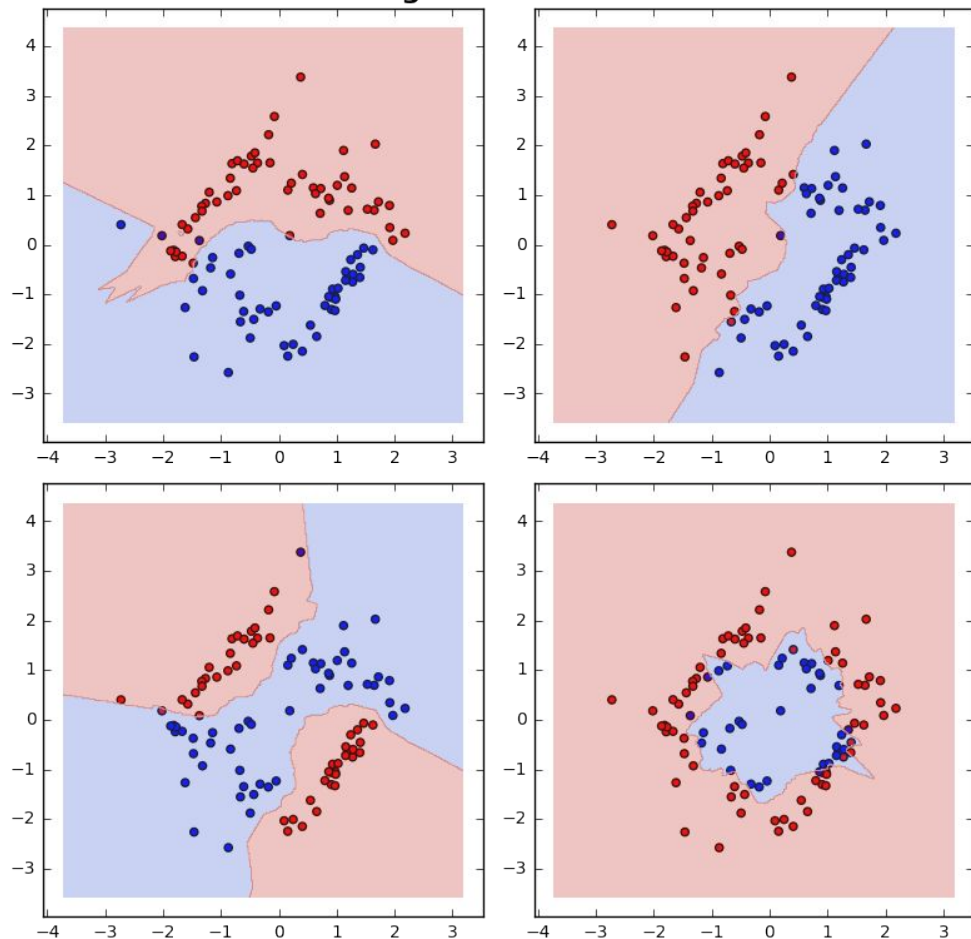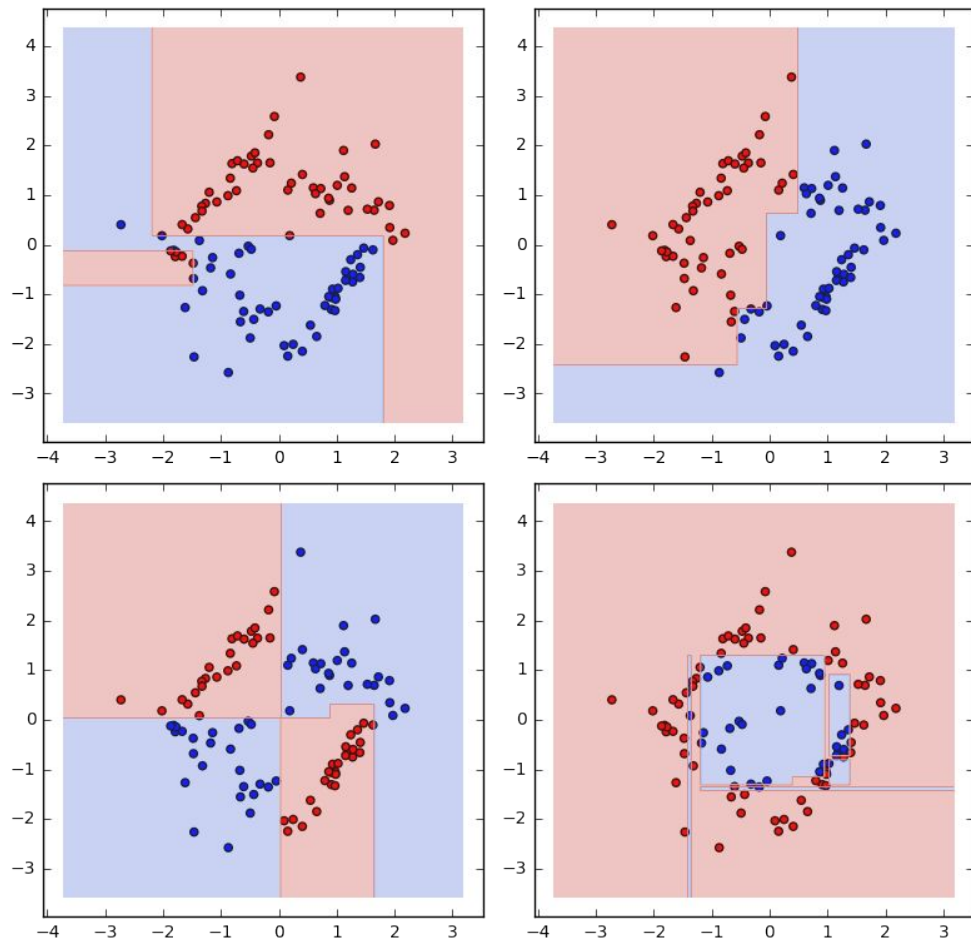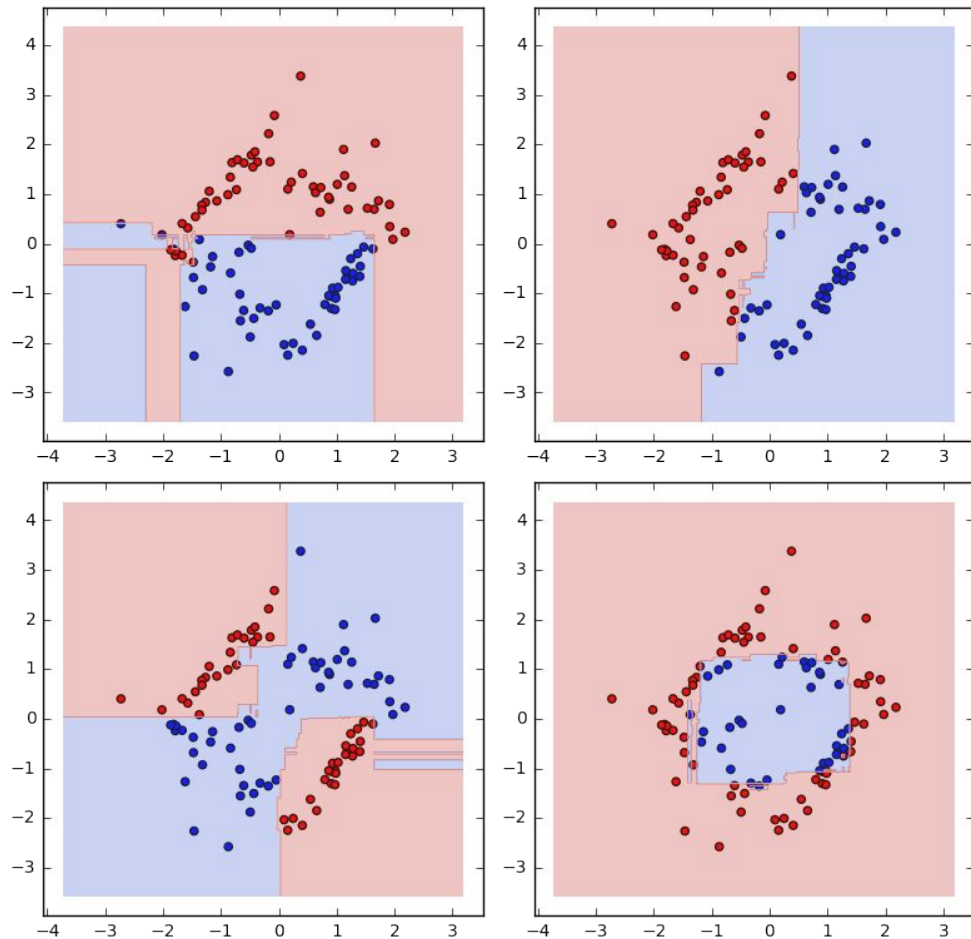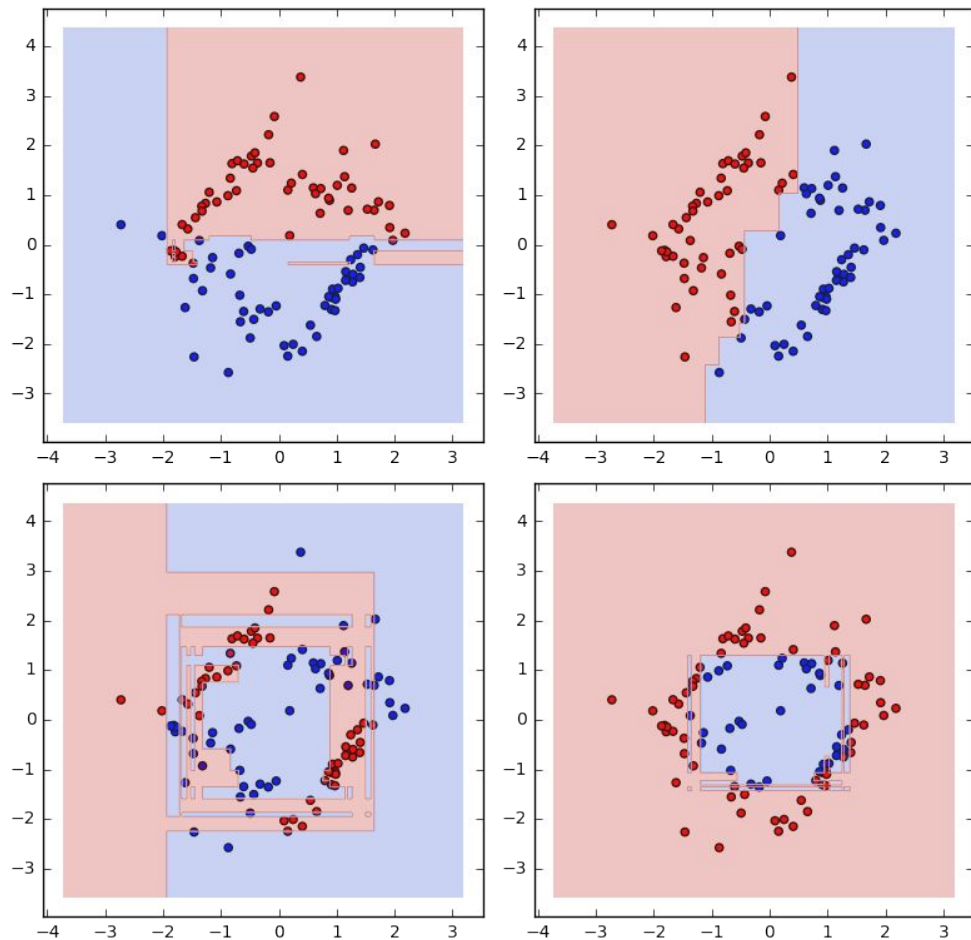
LogisticRegression

KNeighborsClassifier
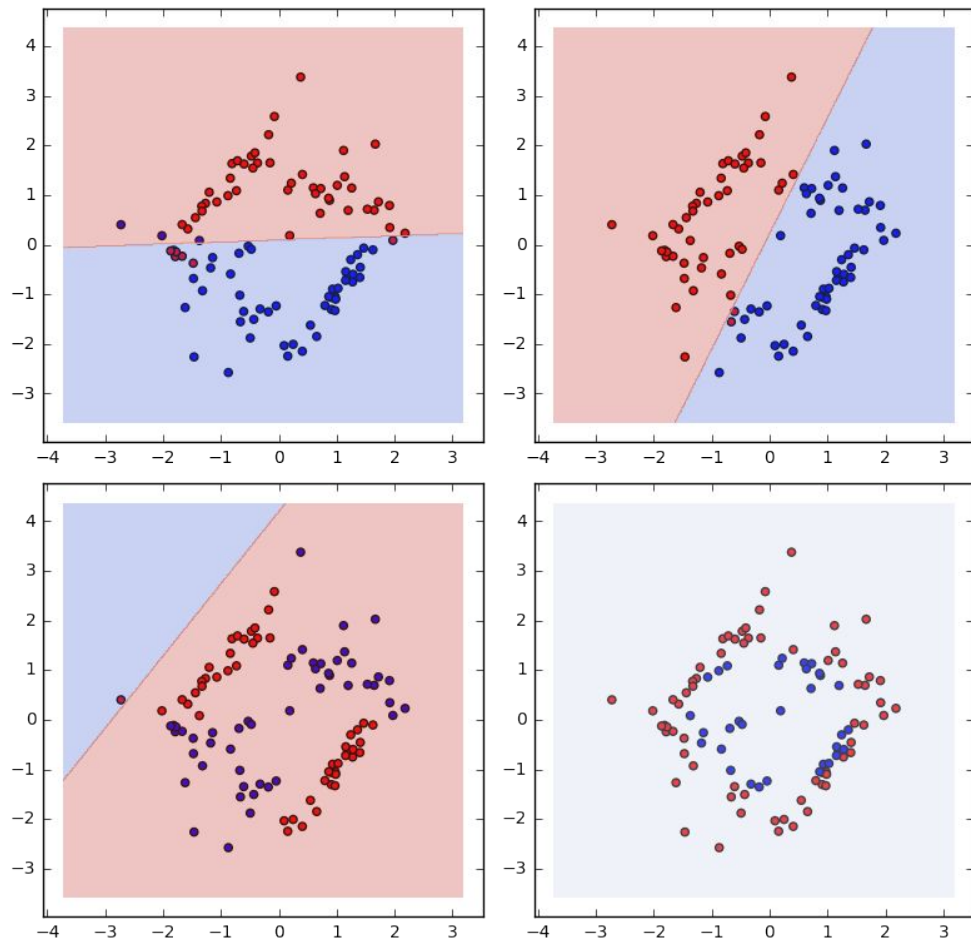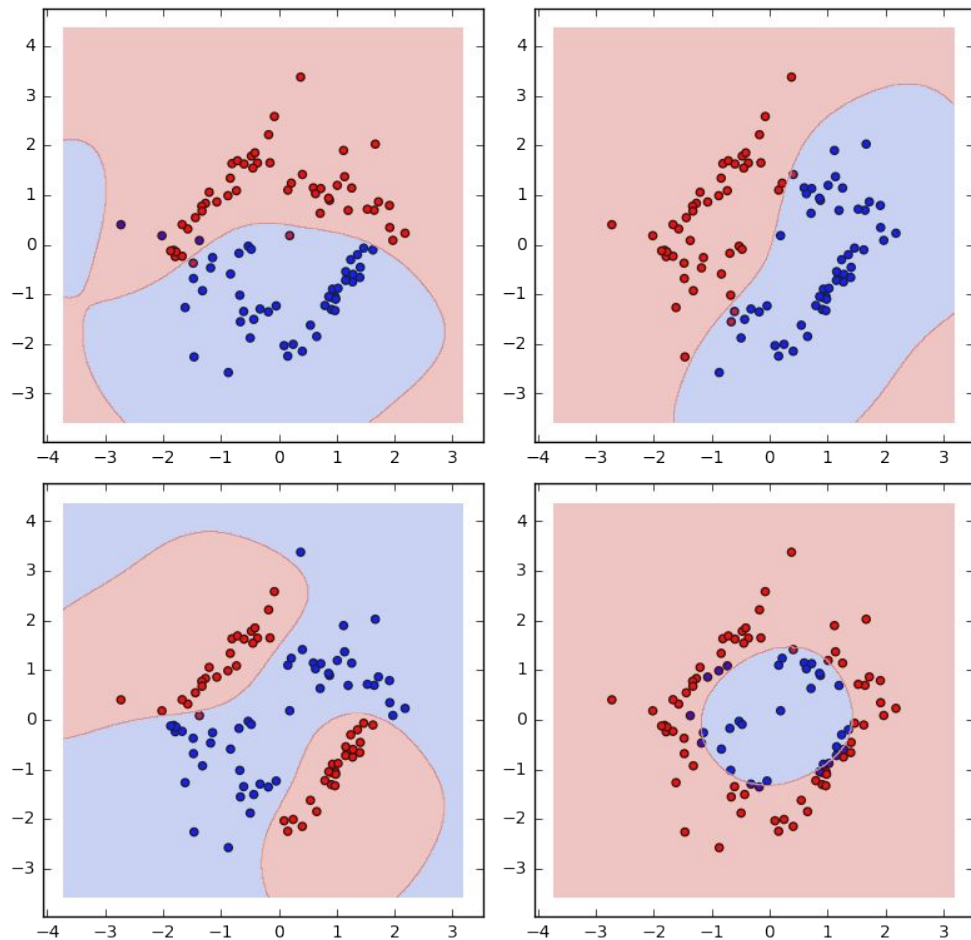
DecisionTreeClassifier
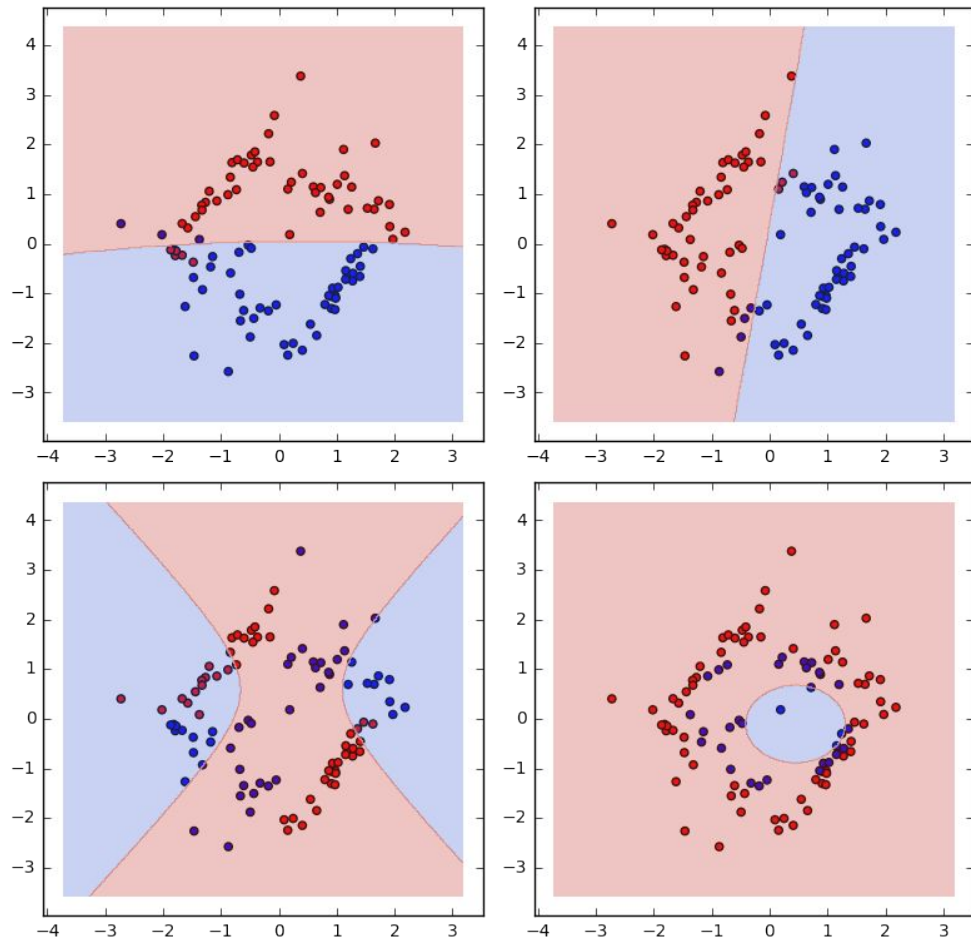
RandomForestClassifier

AdaBoostClassifier

SVC

SVC

GaussianNB

Individual exercise

- What is the purpose of **cross validation**?

Pair discussion

- How do you ***score*** a regression model? A classification model?
- What scores can you use to compare different models? (either different types of models or models built on different feature sets)

Online updating

- Which models can incorporate new data points on-the-fly?
- Which models need to be entirely re-trained in order to incorporate new data?
- How often should I re-train my model if I'm a startup? Amazon? Google?

SUPERVISED LEARNERS

	Regression	Classification	Many dims?	
Lin. Reg.	X		overfitz	
kNN	X	X	The curse!	
Dec. Tree	X	X	Cool with it	
RF	X	X	Cool with it	
Boosted Trees	X	X	Cool with it	
Log. Reg.		X	overfitz	
SVM		X	Protected by a magical amulet	
Neural Net		X	Cool with it	
Naive Bayes		X	V. cool with it.	

UNSUPERVISED LEARNERS

	High dim?			
k-means	The curse!			
hierarchical	The curse!			
NMF				
LDA				