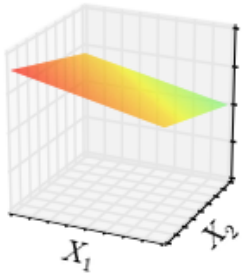
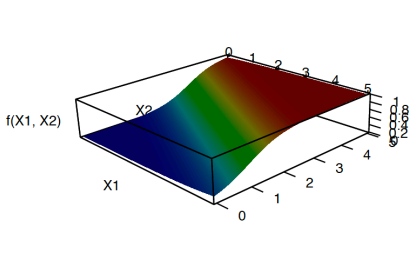
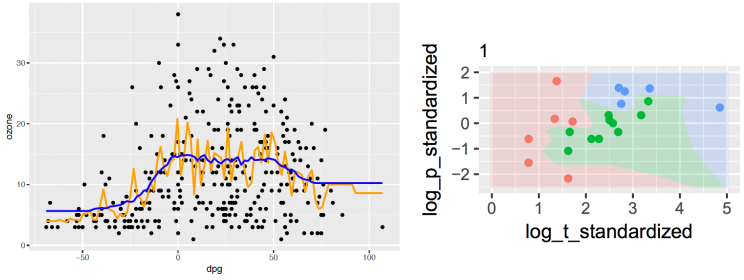
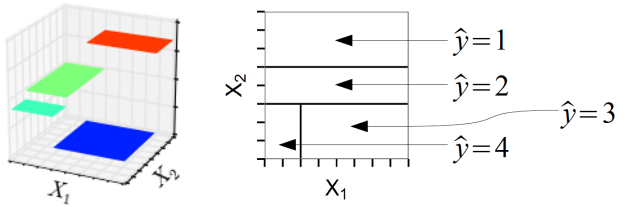
Model	Task	Useful For	Equation	Looks Like
Linear Regression • Least Squares • LASSO • Ridge Regression	Regression	Inference and Prediction (for LASSO and Ridge, Bootstrap or Bayesian approaches to inference)	$\hat{f}(x_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$	
Logistic Regression	Classification	Inference and Prediction	$\hat{f}_1(x_i) = \frac{e^{\beta_0+\beta_1 X_{i1}+\cdots+\beta_p X_{ip}}}{1+e^{\beta_0+\beta_1 X_{i1}+\cdots+\beta_p X_{ip}}}$	
KNN	Regression and Classification	Prediction	$\hat{f}(x_i) = \frac{1}{K}\sum_{i\in N_0} Y_i$ $\hat{f}_j(x_i) = \frac{1}{K}\sum_{i\in N_0} \mathbb{I}_{\{j\}}(Y_i)$	
Classification and Regression Trees	Regression and Classification	Prediction	$\hat{f}(x_i) = \sum_{m=1}^{\lvert T\rvert} I_{R_m}(x_i)\hat{y}_m$ $\hat{y}_m$ is the mean training set response in region m for regression, or the proportion of training set responses in class j from region m.	

Ensembling Approach	Useful For	Component Models Differentiated By…	Timing of Component Model Estimation	Component Models Combined Via…
Stacking	Prediction	Could be anything; often, different types of models (linear model, KNN, CART)	Separate Estimation of Each Component Model, Combined Later	Stage 2 model takes predictions from stage 1 models as inputs. (Stage 2 could be a simple average or majority vote, or a complex model).
Boosting	Prediction	Each component model trained on residuals from previous models, or more weight assigned to cases not predicted well by previous models.	Sequential Estimation of Component Models, Estimation of Each Model Depends on Previous Models	Typically, additive (predictions from component models are added together)

Approach for Model Differentiation	Component Models Differentiated By…
Bagging	Models trained on bootstrap resampled data sets
Feature Subsets	Models use different subsets of features, or use features differently in estimation process

Named Ensemble Model	Component Models Differentiated By…
Random Forest	All component models are trees. Combines bagging and feature subsets: each tree is trained on a bagged data set, and a randomly selected subset of features are used in finding each split as the tree is grown.