* Review from last time – bias variance tradeoff
  + The variance is high for the cubic
  + Can decompose into bias and variance
  + Cubic fits better but not enough training data
* Recap complete pipeline
  + Training data, then pick a model class and loss function, perform cross validation and maybe choose a model
  + Then profit
* Today’s lecture: perceptron algorithm and stochastic gradient descent
* Perceptron
  + Going to be a linear classification model so it’ll give a sign
  + For each step we receive a random training example then we make a prediction and see if current version of model matches the label
  + If
* Linear models computer un normalized signed distance
* There are many possible solutions for the perceptron
* Convergence to a mistake free implies that it’s linearly separable
* When gamma is small there’s fewer possible solutions
* The number of mistakes is bounded by the radius of the feature space and the margin gamma
* But most problems are not linearly separable!
  + Use validation set and do early stopping
* Online learning – get a stream of data one piece at a time
* Can convert batch to online learning
* If the step size is too big then you’ll oscillate infinitely and small steps take too long
* Want the step size as large as possible without diverging
* Mostly trial and error picking step size
* Takes infinite time, but we don’t care because it just has to be good enough
* Could stop when progress is sufficiently small