

Ensembles: Combining classifiers

- Can a set of weak classifiers be combined to derive a strong classifier? **Yes!**
- Average results from different models
- Why?
 - Better classification performance than individual classifiers
 - More resilience to noise
- Why not?
 - Time consuming
 - Models become difficult to explain

“Wisdom of the the (simulated) crowd”

Bagging

- Draw N bootstrap samples
- Retrain the model on each sample
- Average the results
 - Regression: Averaging
 - Classification: Majority vote
- Works great for overfit models
 - Decreases variance without changing bias
 - Doesn't help much with underfit/high bias models
 - Insensitive to the training data

Boosting

- Instead of selecting data points randomly with the bootstrap, **favor the misclassified points**

Initialize the weights

Repeat:

 Resample with respect to weights

 Retrain the model

 Recompute weights

For each step t

$D_t(i)$ weights: probability of selecting
example i in the sample

x_i, y_i i th example, i th label

h_t trained classifier at step t using sample drawn
according to D_t

$\epsilon_t = \sum_{i: h_t(x_i) \neq y_i} D_t(i)$ sum of weights for
misclassified examples

$\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}$ odds of misclassifying

$D_{t+1}(i) = \beta_t D_t(i)$ adjust weights down for
correctly classified examples

...and normalize to make
sure weights sum to 1