

Regularization Comparison

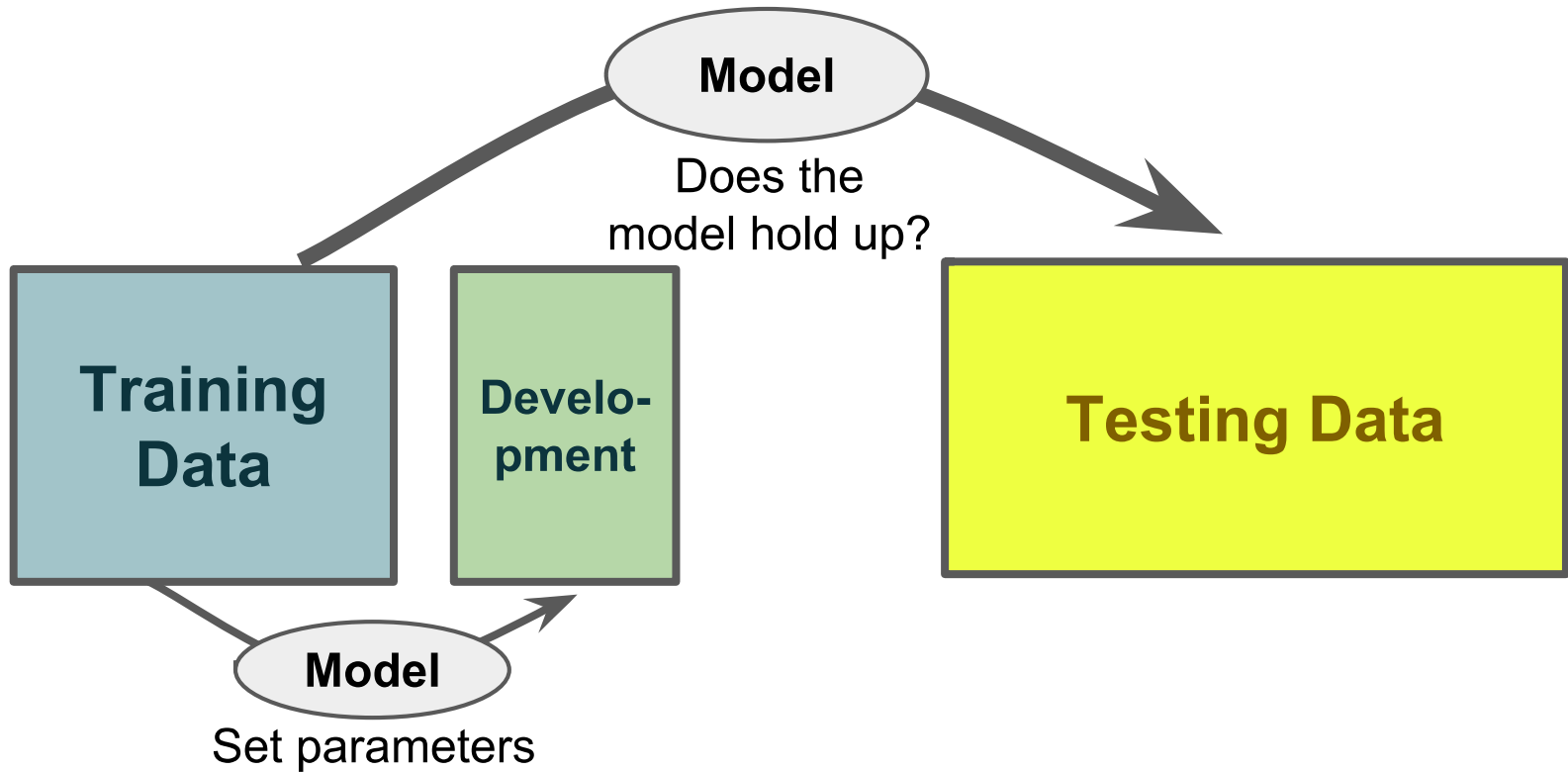


Review, 3/31 - 4/5

- Confidence intervals
- Bootstrap

- Prediction Framework: Train, Development, Test
- Overfitting: Bias versus Variance
- Feature Selection: Forward Stepwise Regression
- Ridge Regression (L2 regularization)
- Lasso Regression (L1 regularization)

Common Goal: Generalize to new data



N-Fold Cross-Validation

Goal: Decent estimate of model accuracy



Fold 1



Fold 2



Fold 3



....

...

Supervised vs. Unsupervised

Supervised

- Predicting an outcome $E(y|X)$
- Loss function used to characterize quality of prediction

$$L(y, \hat{y}) = (y - \hat{y})^2$$

Supervised vs. Unsupervised

Supervised

- Predicting an outcome $E(y|X)$
- Loss function used to characterize quality of prediction

$$L(y, \hat{y}) = (y - \hat{y})^2$$

Unsupervised

- No outcome to predict
- Goal: Infer properties of $P(X)$ without a supervised loss function.
- Often larger data.
- Don't need to worry about conditioning on another variable.

K-Means Clustering

Clustering: Group similar observations, often over unlabeled data.

K-means: A “prototype” method
(i.e. not based on an algebraic model).

Euclidean Distance:
$$d(x_i, x_{i'}) = \sum_{j=1}^m (x_{ij} - x_{i'j})^2 = \|x_i - x_{i'}\|^2$$

centers = a random selection of k cluster centers

until centers converge:

1. For all x_i , find the closest center (according to d)
2. Recalculate centers based on mean of euclidean distance

Dimensionality Reduction - Concept

