ML & Advanced Analytics For Biomedicine

Ishanu Chattopadhyay
Department of Medicine
University of Chicago

CCTS 40500 / CCTS 20500 / BIOS 29208

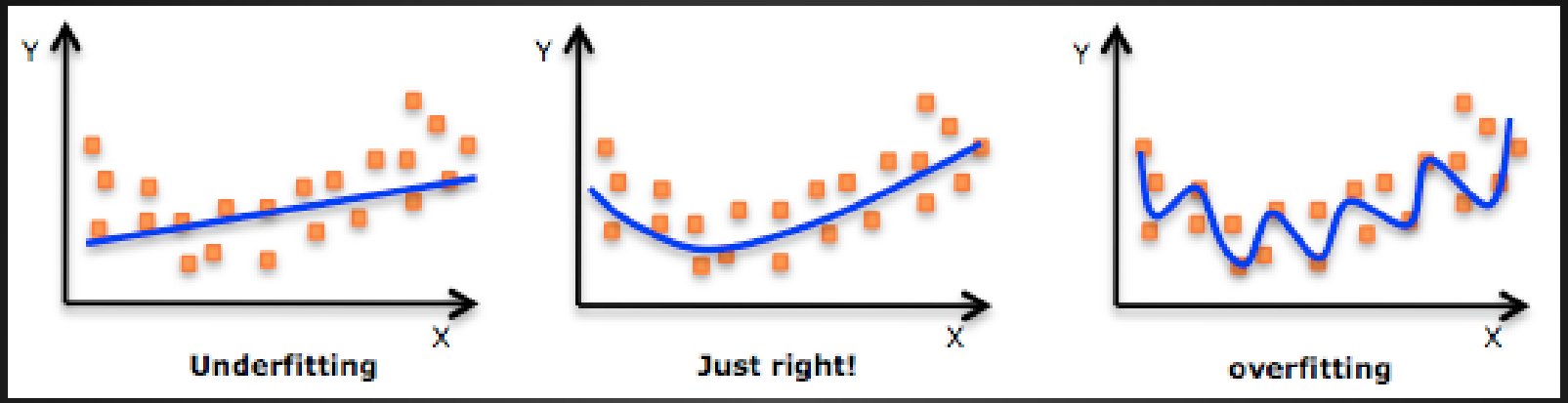
Winter 2019

Today

- **Bagging & Boosting**
- **Regressors**
- **Bias Vs Variance**

Regression

Fitting, Overfitting & The Sweet Spot



- **How complex should be our model?**
- **Bias Vs Variance**
- **Regularization**
  - Ridge (ℓ_2 regularization)
  - Lasso (ℓ_1 regularization)
  - Elastic Net (convex combo of ridge and LASSO)

Regression Algorithms

Data: $y_i \in \mathbb{R}$

Model: $y_i = w^T x_i + \epsilon_i$ **where** $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

$$w = \arg\min_w \frac{1}{n} \sum_{i=1}^{n} (x_i^T w - y_i)^2 \qquad \text{(Linear)}$$

$$w = \arg\min_w \frac{1}{n} \sum_{i=1}^{n} (x_i^T w - y_i)^2 + \lambda ||w||_2^2 \qquad \text{(Ridge)}$$

$$w = \arg\min_w \frac{1}{n} \sum_{i=1}^{n} (x_i^T w - y_i)^2 + \lambda ||w||_1 \qquad \text{(LASSO)}$$

Regression Algorithms

Performance Metrics

- **R Squared**
- **Adjusted R Squared**
- **Pearson's Rho**

Bias Variance Decomposition

Error = Bias + Variance + Noise
Bagging reduces variance
Boosting reduces bias