INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

Lecture 12

Hassan Ashtiani

EXPECTED ERROR MINIMIZATION

- The solution to $\boxed{y^*} = argmin_{\hat{y}}\, E_{X,Y}\left(Y - \hat{y}(X)\right)^2$
  - Is $y^*(x) = E[Y|X = x]$
- In practice, we don't know $P(X, Y)$
  - Cannot calculate $y^*(x)$
- Instead, we have $Z = \left\{\left(x^i, y^i\right)\right\}_{i=1}^n$
  - We have used Z to find predictor $\widehat{y_Z}$
- What can we say about $E_{x,y}\left(y - \widehat{y_Z}(x)\right)^2$?

ERROR DECOMPOSITION

- $E_{x,y}\left(y - \widehat{y_Z}(x)\right)^2$

$$= E_{x,y}\left(y - y^*(x)\right)^2 + E_x\left(y^*(x) - \widehat{y_Z}(x)\right)^2$$

$$= \textit{noise} + \text{ESTIMATION ERROR}$$

does not depend on $\hat{y}$

$$\hat{y}_2$$

$$\boxed{y^*(x) = E[y|x]}$$

$$\underset{(x,y)\sim p(x,y)}{E} \left(y - \hat{y}_2(x)\right)^2 =$$

$$= \underset{(x,y)}{E} \left(y - y^*(x) + y^*(x) - \hat{y}_2(x)\right)$$

$$= \underbrace{\underset{x,y}{E} \left(y - y^*(x)\right)^2}_{\text{noise}} + \underbrace{E\left(\hat{y}_2(x) - y^*(x)\right)^2}_{\text{estimation error}} + \underbrace{2E\left(y - y^*(x)\right)\left(y^*(x) - \hat{y}_2(x)\right)}_{\text{cross term} = 0}$$

crossterm: $$2 \underset{x\sim p(x)}{E} \left(\underset{y\sim p(y|x)}{E} \left[y \cdot y^*(x) + y^*(x)\hat{y}_2(x) - y\hat{y}_2(x) - y^*(x)y^*(x) \right] \right)$$

$$= 2 \underset{x\sim p(x)}{E} \left[\cancel{y^*(x)y^*(x)} + \cancel{y^*(x)\hat{y}_2(x)} - \cancel{y^*(x)\hat{y}_2(x)} - \cancel{y^*(x)y^*(x)} \right]$$

BIAS-VARIANCE DECOMPOSITION

- DOES $E_{x,y}\left(y - \widehat{y_Z}(x)\right)^2$ CAPTURE THE QUALITY OF OUR "ESTIMATION METHOD"?
  - DEPENDS ON Z….
- HOW GOOD IS AN ESTIMATOR "ON AVERAGE"?

$E\ Z$

$$E_Z E_{x,y}\left(y - \widehat{y_Z}(x)\right)^2 = ?$$

E.G., FOR LINEAR ESTIMATOR:

$$E_Z E_{x,y}\left(y - W_Z^T x\right)^2 = ?$$

BIAS-VARIANCE DECOMPOSITION

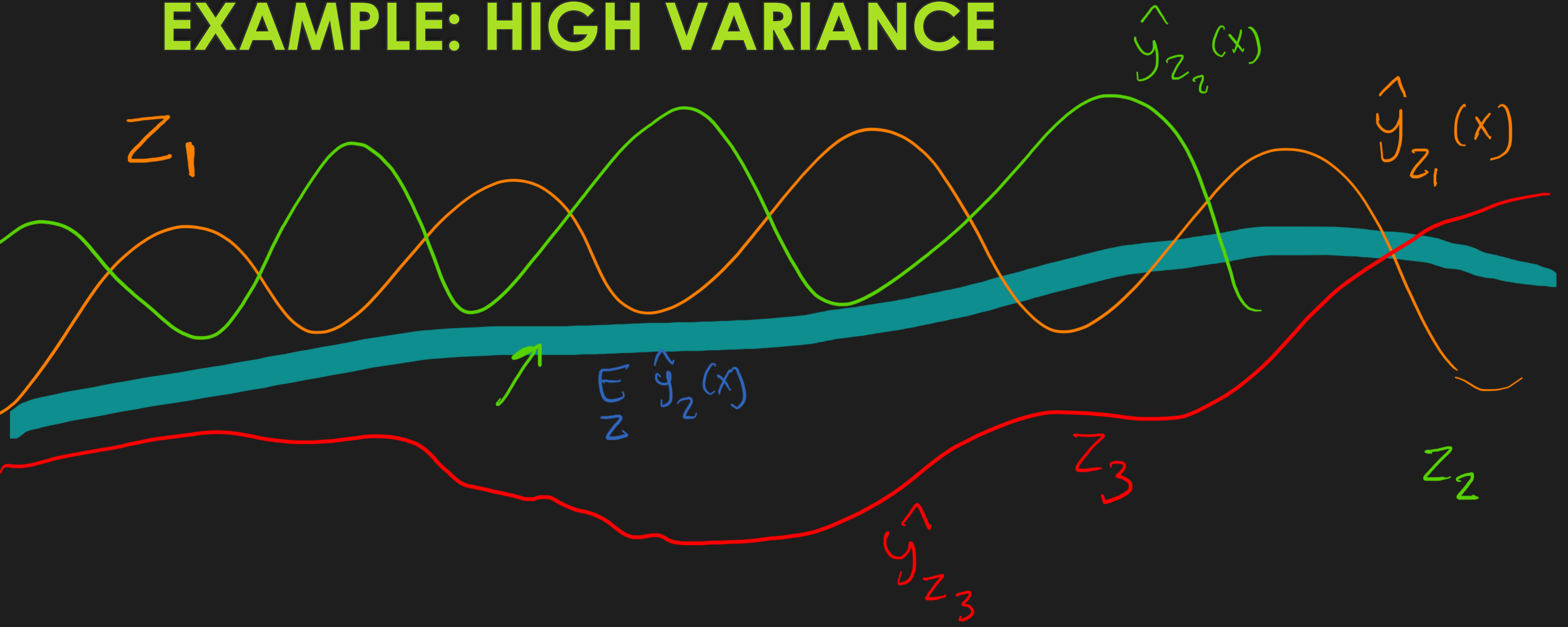
$$E_Z E_{x,y}(y - \widehat{y_Z}(x))^2$$

$$= E_{x,y}(y - y^*(x))^2 \quad \textbf{(NOISE)}$$

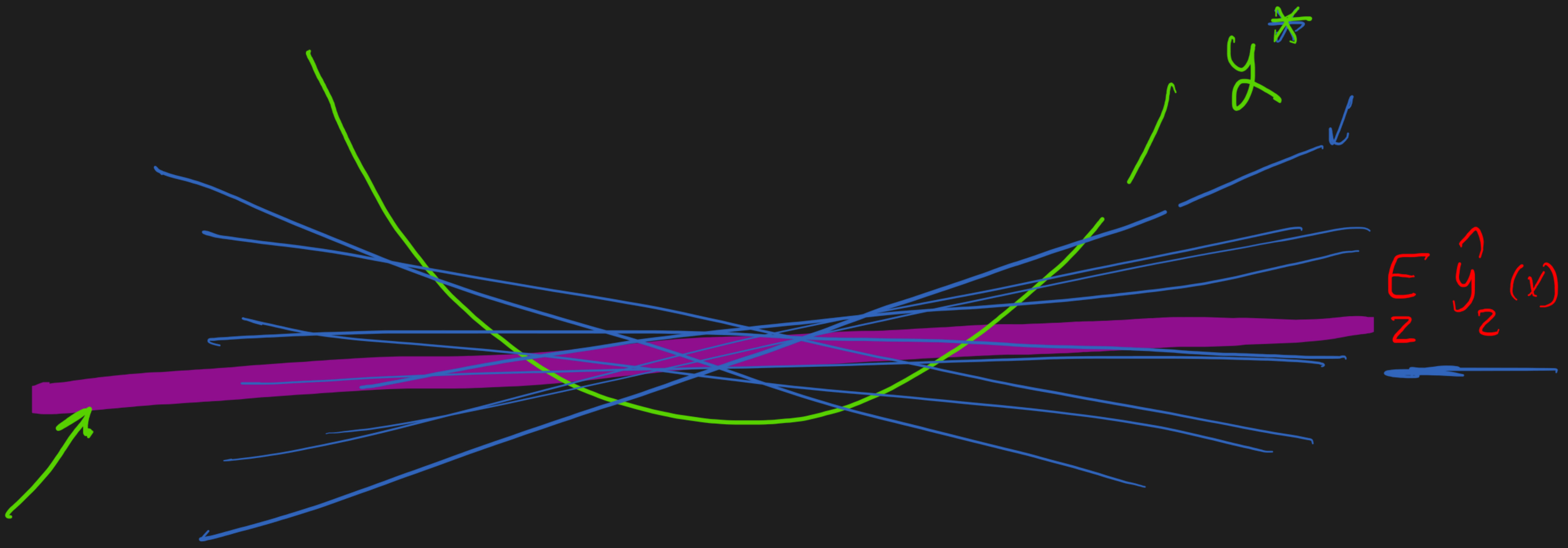
$$+E_x(y^*(x) - E_Z(\widehat{y_Z}(x)))^2 \quad \textbf{(BIAS)}$$

$$+E_x E_Z(\widehat{y_Z}(x) - E_Z(\widehat{y_Z}(x)))^2 \quad \textbf{(VARIANCE)}$$

EXAMPLE: HIGH VARIANCE



EXAMPLE: HIGH BIAS , low variance



Low bias, high variance?