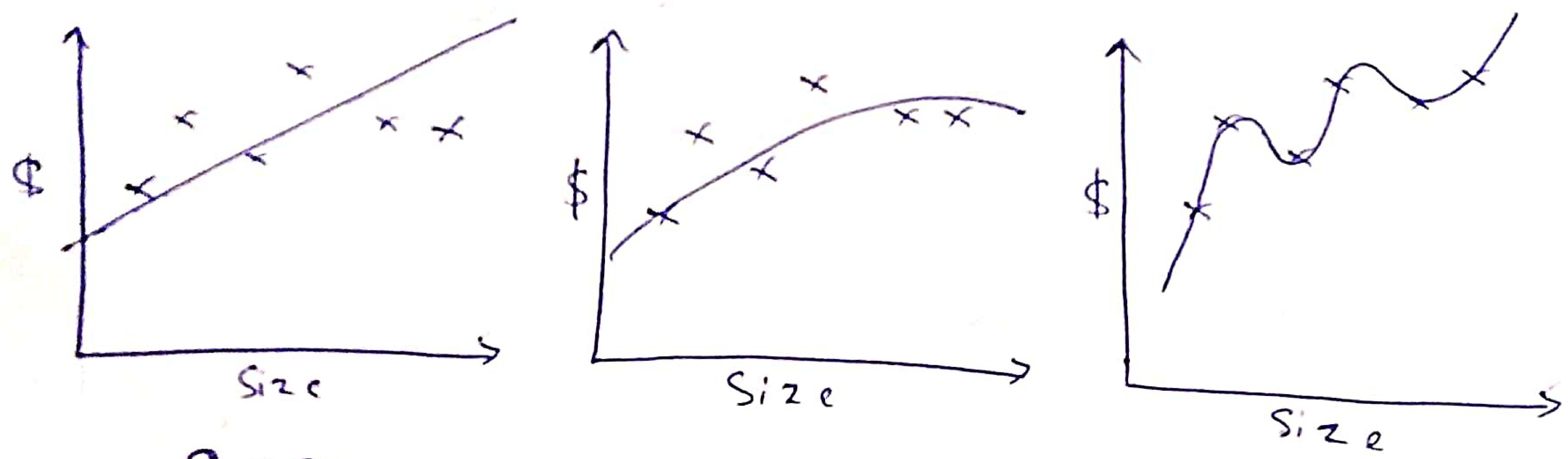
8

Data Split, Models & Cross-Validation



$\theta_0 + \theta_1 x$ | $\theta_0 + \theta_1 x + \theta_2 x^2$ | $\theta_0 + \theta_1 x + \cdots + \theta_5 x^5$

Underfit | Just right | Overfit

high bias | | high variance

{Very Strong Preconception}

{Changing training dataset little bit will drastikly change the fitted curve}

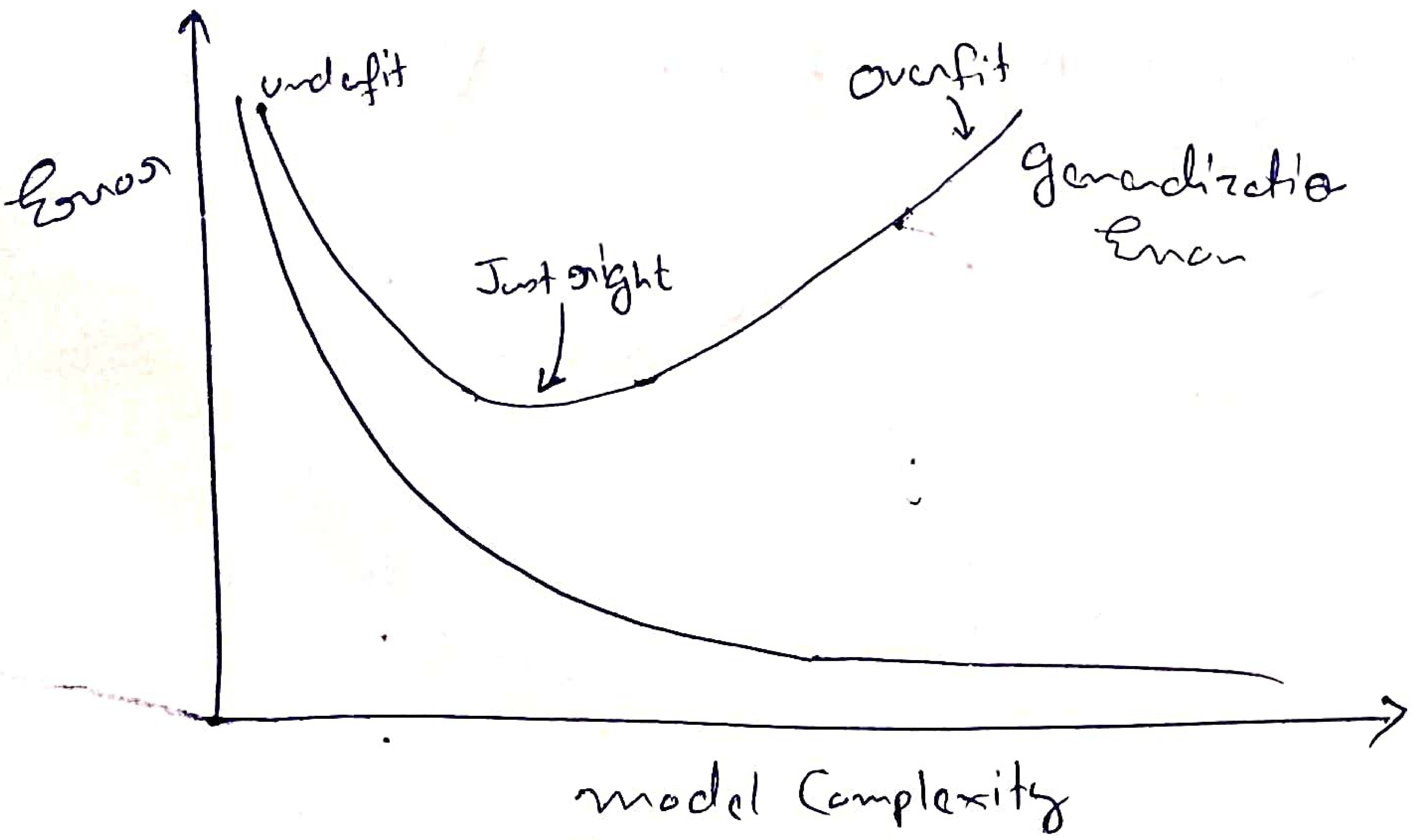
⇒ One of the most effective ways to prevent overfitting is regularization.

★ **Regularization**

$$\min_{\theta} \frac{1}{2} \sum_{i=1}^{m} \| y^{(i)} - \theta^T x^{(i)} \|^2$$ {Cost function for Linear regression}

After regularization

$$\min_{\theta} \frac{1}{2} \sum_{i=1}^{m} \| y^{(i)} - \theta^T x^{(i)} \|^2 + \frac{\lambda}{2} \|\theta\|^2$$



★ **Simple Hold out Cross validation**

(Train Set) + (~~dev set~~ Dev Set) + (Test Set) = (Data Set)

S_{train} S_{dev} S_{test} S

60%   20%   20%

{Also called Cross-validation Set}

⇒ Train each model on S_{train}

↳ Get some hypothesis h_i

⇒ Measure the error on S_{dev}. Pick the one with lowest error on S_{dev}.

⇒ Evaluate the algorithm on separate S_{test} and report that error.

★ K-fold CV {Use this only for small dataset}

$$\begin{bmatrix} x^{(1)} & y^{(1)} \\ \vdots & \\ x^{(100)} & y^{(100)} \end{bmatrix}$$

⇒ Divide dataset into ~~[illegible]~~ K subsets. (let $K=5$)

→ For $i = 1 \ldots K$

→ Train on $K-1$ pieces

→ Test on the remaining 1 piece

→ Take Average.

→ Refit the best model found on 100% of the data

★ Feature Selection

Start with $F = \phi$ ← Forward Search

Repeat

1) Try adding each feature to F, and see which feature addition most improves the dev set performance.

2) Add that feature to F.