REGRESSION

Lecture 3

MAL1, 2025

REGRESSION

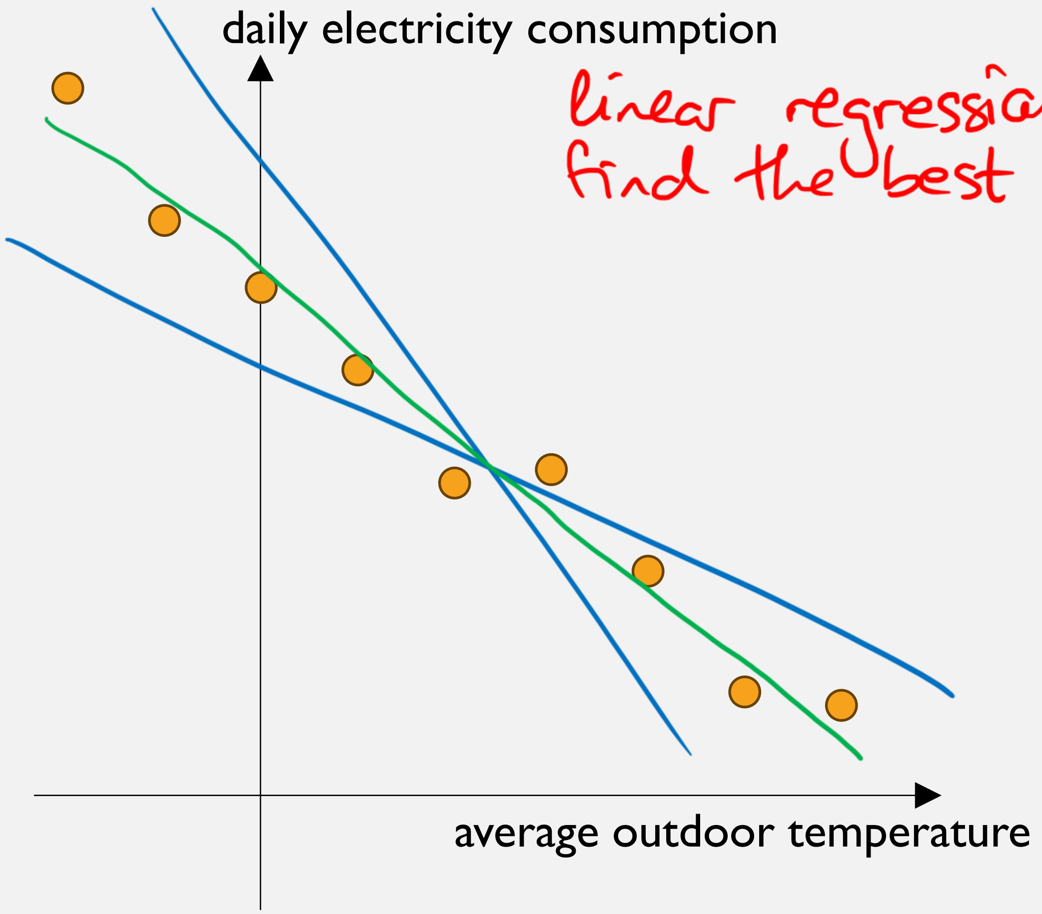
- Linear regression
- Performance metrics
- Polynomial regression
- Regularization

REGRESSION

average outdoor temperature (°C)	daily electricity consumption (kWh)
-10	46.5
-5	37.9
0	33.2
5	27.5
10	20.3
15	21.1
20	14.2
25	6.3
30	5.6

feature

response variable

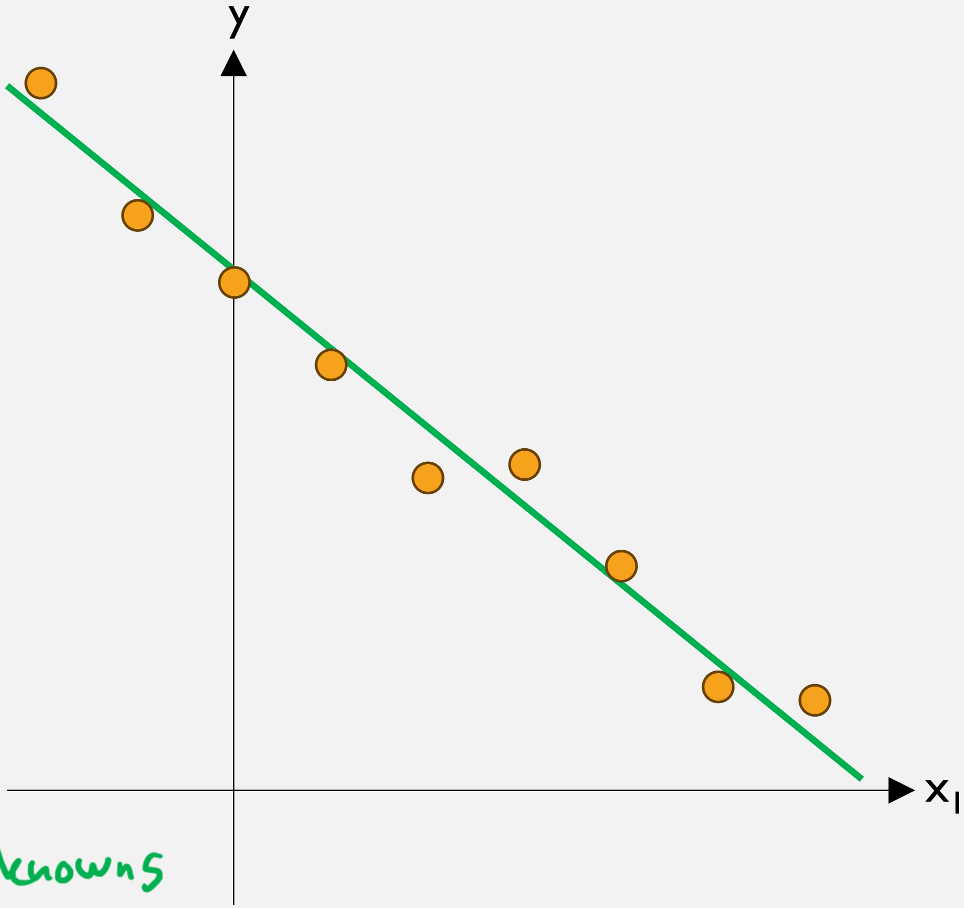


linear regression: find the best line

REGRESSION

x_1	y
β_0 + -10 $\times \beta_1$	$\approx$ 46.5
β_0 + -5 $\times \beta_1$	$\approx$ 37.9
β_0 + 0 $\times \beta_1$	$\approx$ 33.2
β_0 + 5 $\times \beta_1$	$\approx$ 27.5
β_0 + 10 $\times \beta_1$	$\approx$ 20.3
β_0 + 15 $\times \beta_1$	$\approx$ 21.1
β_0 + 20 $\times \beta_1$	$\approx$ 14.2
β_0 + 25 $\times \beta_1$	$\approx$ 6.3
β_0 + 30 $\times \beta_1$	$\approx$ 5.6

9 equations, 2 unknowns



with $\hat{y} = \beta_0 + \beta_1 x_1$

find the "best" values of β_0 and β_1

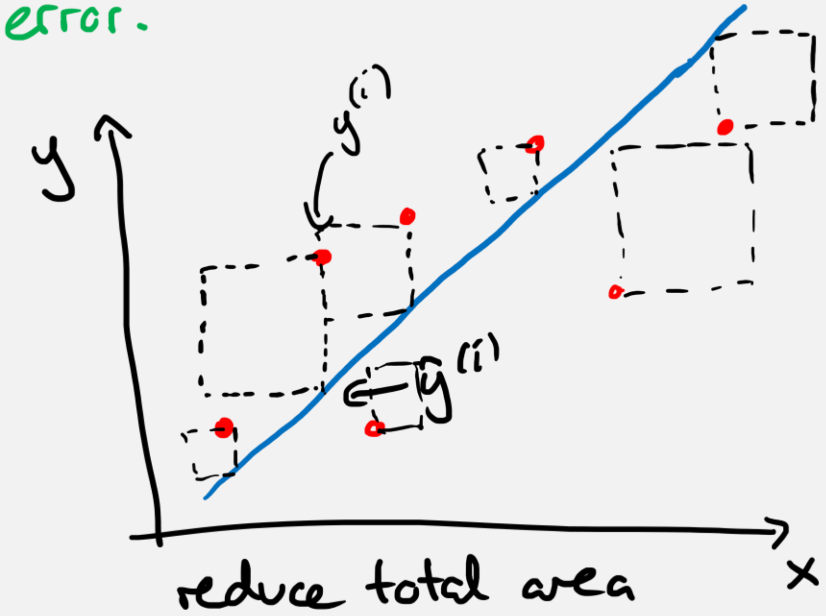
$$\underbrace{\beta_0 = 33.72 \qquad \beta_1 = -1.009}_{\text{finding these numbers = training the model}}$$

FINDING BETA'S

$$\underbrace{\hat{y}}_{\text{prediction}} = \beta_0 + \beta_1 \underbrace{x_1 + \beta_2 x_2 + \beta_3 x_3}_{\text{features}} + \dots = \underbrace{B^T x}_{\text{matrix form}}$$

The best line reduces the error:

$$\underbrace{SSE}_{\text{"sum of squared errors"}} = \sum_i \left(y^{(i)} - \hat{y}^{(i)}\right)^2$$



FINDING BETA'S

$$\text{SSE} = \sum_i \left(y^{(i)} - \hat{y}^{(i)}\right)^2 = \sum_i \left(y^{(i)} - B^T x^{(i)}\right)^2$$

minimize → take the derivative wrt. all β's and set equal to zero:

$$\frac{\partial}{\partial \beta_j} \text{SSE} = \frac{\partial}{\partial \beta_j} \sum_i \left(y^{(i)} - B^T x^{(i)}\right)^2 = 2 \sum_i \left(y^{(i)} - B^T x^{(i)}\right) x_j^{(i)} = 0$$

summarize in matrix form:

$2 \cdot 0 = 0$

$$2\overbrace{\boldsymbol{X}^T(\boldsymbol{XB} - \boldsymbol{y})}^{=0} = 0$$

$$\Rightarrow X^T X B - X^T y = 0$$

$$\Rightarrow X^T X B = X^T y$$

$$\Rightarrow B = (X^T X)^{-1} X^T y$$ ← "normal equation"

FINDING BETA'S

$\hat{y} = \beta_0 \cdot 1 + \beta_1 \cdot x_1$

$$\boldsymbol{B} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & -10 \\ 1 & -5 \\ 1 & 0 \\ 1 & 5 \\ 1 & 10 \\ 1 & 15 \\ 1 & 20 \\ 1 & 25 \\ 1 & 30 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 46.5 \\ 37.9 \\ 33.2 \\ 27.5 \\ 20.3 \\ 21.1 \\ 14.2 \\ 6.3 \\ 5.6 \end{bmatrix}$$

$$\boldsymbol{X}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -10 & -5 & 0 & 5 & 10 & 15 & 20 & 25 & 30 \end{bmatrix}$$

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 9 & 90 \\ 90 & 2400 \end{bmatrix} \qquad (\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} \frac{8}{45} & -\frac{1}{150} \\ -\frac{1}{150} & \frac{1}{1500} \end{bmatrix}$$

$$\boldsymbol{X}^T\boldsymbol{y} = \begin{bmatrix} 212.6 \\ 612 \end{bmatrix}$$

$$\boldsymbol{B} = \overbrace{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T}^{\text{pseudoinverse}}\boldsymbol{y} = \begin{bmatrix} 33.72 \\ -1.009 \end{bmatrix}$$

β_0 intercept

β_1 slope

new variable matching β_0

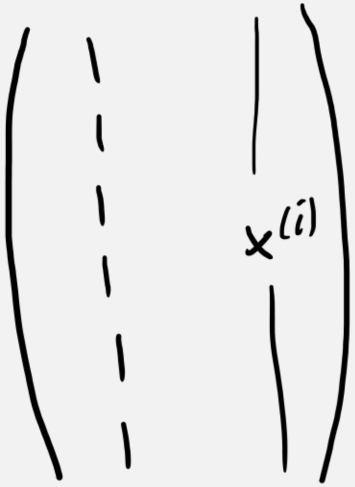
x_0	x_1	y
1	-10	46.5
1	-5	37.9
1	0	33.2
1	5	27.5
1	10	20.3
1	15	21.1
1	20	14.2
1	25	6.3
1	30	5.6

↳ design matrix

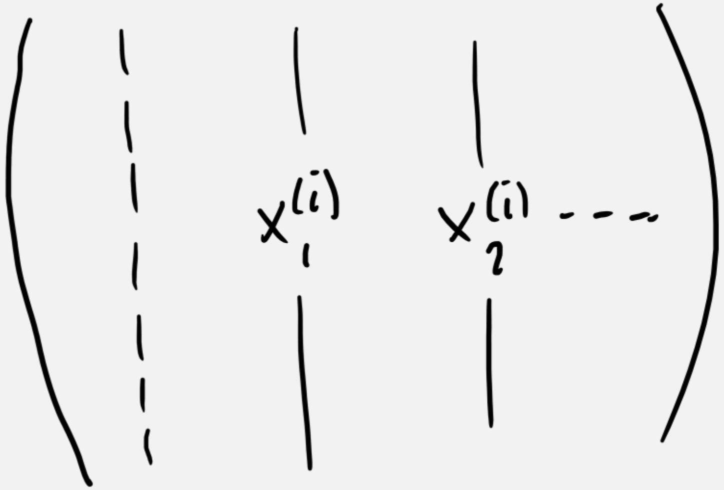
in practice, the pseudoinverse is computed using SVD

THE DESIGN MATRIX

One variable



Multiple variables



REGRESSION

- Linear regression
- Performance metrics
- Polynomial regression
- Regularization

SSE AND FRIENDS

- The smaller the SSE, the better the model …

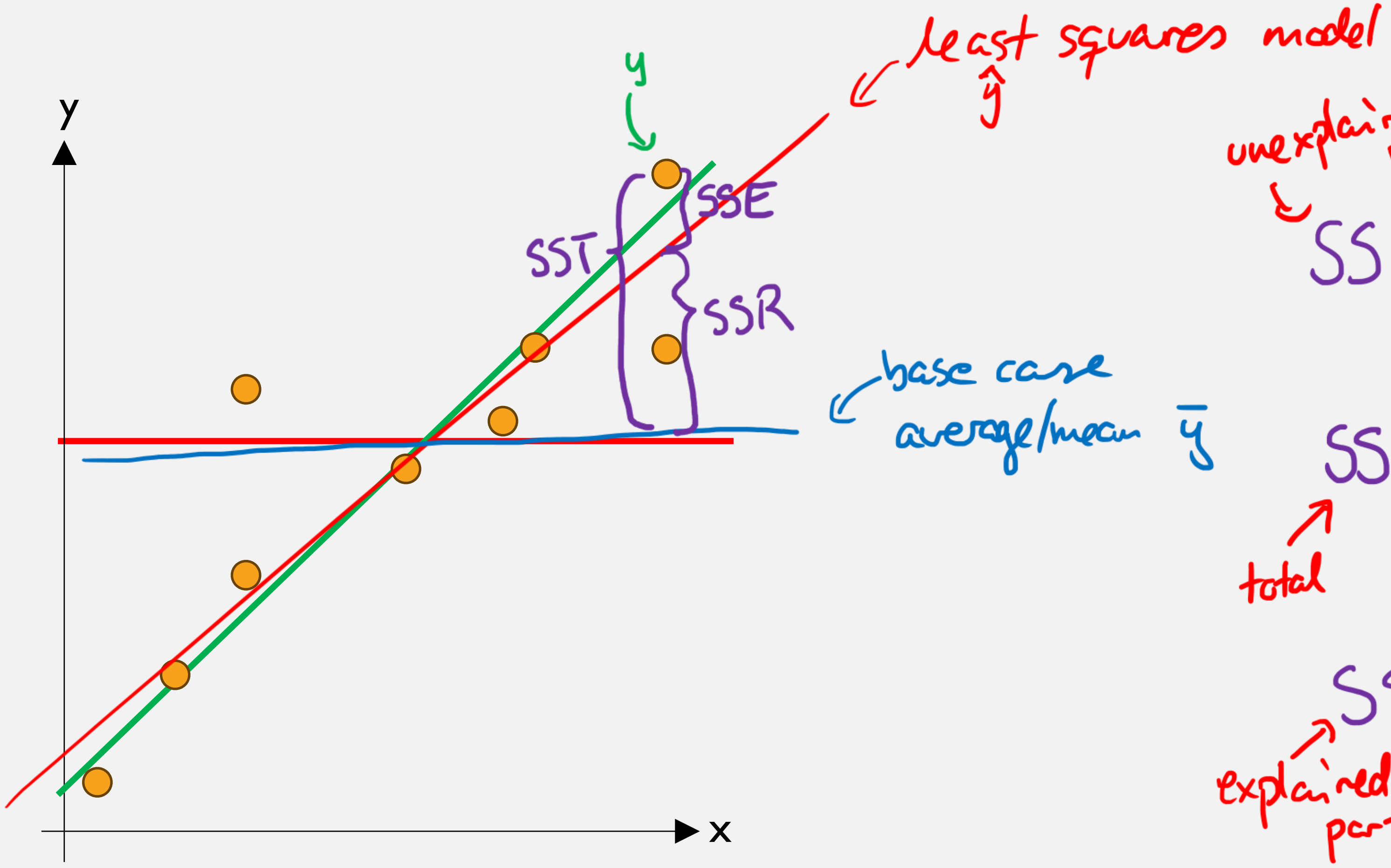
but SSE depends on # data points

$$\rightarrow \quad MSE = \frac{1}{n} SSE$$ (mean squared error)

but MSE depends on the scale of response variable

… so what should we use as our performance metric?

SSE AND FRIENDS



unexplained part

$$SSE = \sum_i \left(y^{(i)} - \hat{y}^{(i)}\right)^2$$

$$SST = \sum_i \left(y^{(i)} - \bar{y}\right)^2$$

total

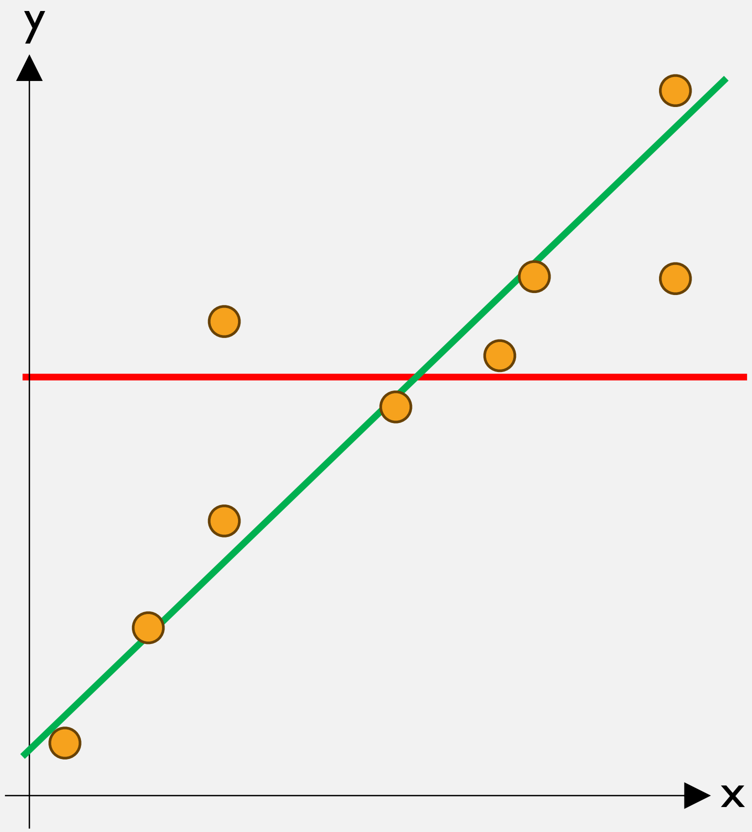
(total sum of squares)

$$SSR = \sum_i \left(\hat{y}^{(i)} - \bar{y}\right)^2$$

explained part

(regression sum of squares

SSE AND FRIENDS



$$\text{SST} = \sum_i \left(y^{(i)} - \bar{y}\right)^2$$ total deviation from mean

$$\text{SSE} = \sum_i \left(y^{(i)} - \hat{y}^{(i)}\right)^2$$ unexplained part

$$\text{SSR} = \sum_i \left(\hat{y}^{(i)} - \bar{y}\right)^2$$ explained part

performane metric : r^2

$$r^2 = \frac{SSR}{SST} = \frac{\text{"explained"}}{\text{"total"}} \leq 1$$

=1 means perfect predictive model

the amount of variance is the model able to explain

CODE EXAMPLE

Jupyter Notebook **Regression - Hitters**

REGRESSION

- Linear regression
- Performance metrics
- Polynomial regression
- Regularization

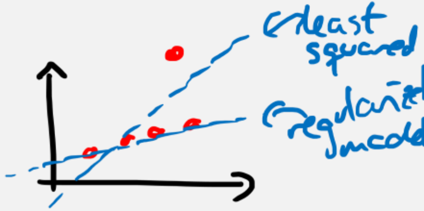
POLYNOMIAL REGRESSION

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots$$

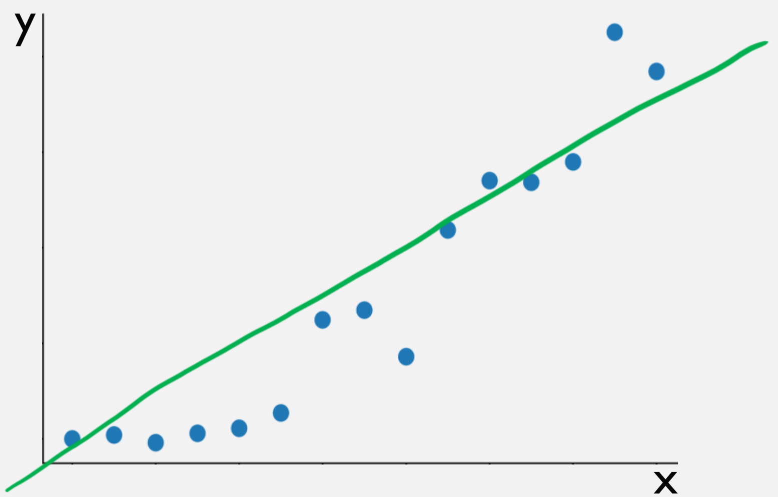
x^2, x^3 are just new features

$$X = \begin{bmatrix} \vdots & | & | & | & \\ \vdots & x & x^2 & x^3 & \cdots \\ \vdots & | & | & | & \end{bmatrix}$$

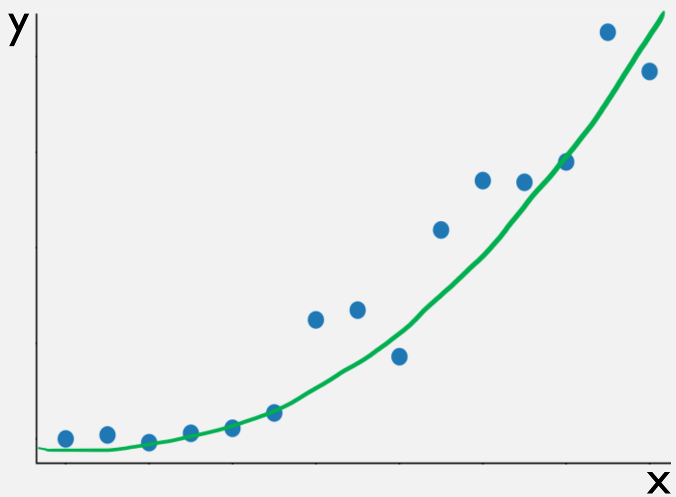
may be a good idea



UNDERFITTING AND OVERFITTING

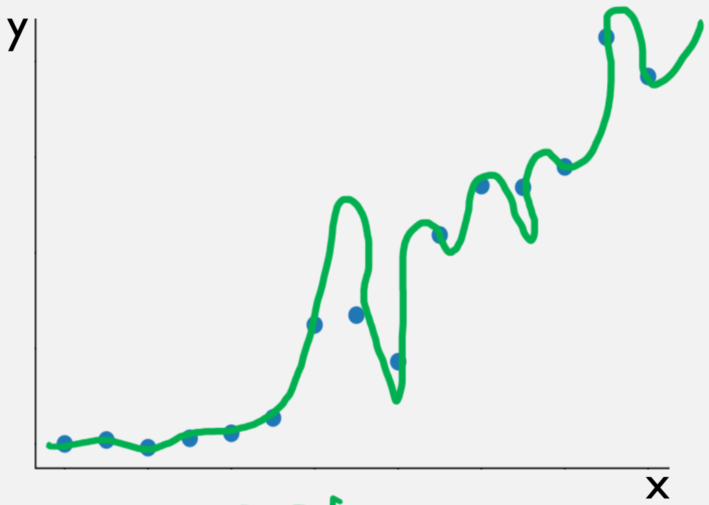


underfitting
High bias
Low variance



sweet spot

bias-variance trade-off



overfitting
Low biased
High variance

Bias : Inability to learn from data
Variance : Reliance on data

REGRESSION

- Linear regression
- Performance metrics
- Polynomial regression
- Regularization

REGULARIZATION

Tool to avoid overfitting

Idea: Penalize large coefficients

$$\text{loss function} \quad \mathcal{L} = MSE + \alpha \cdot R(\beta)$$

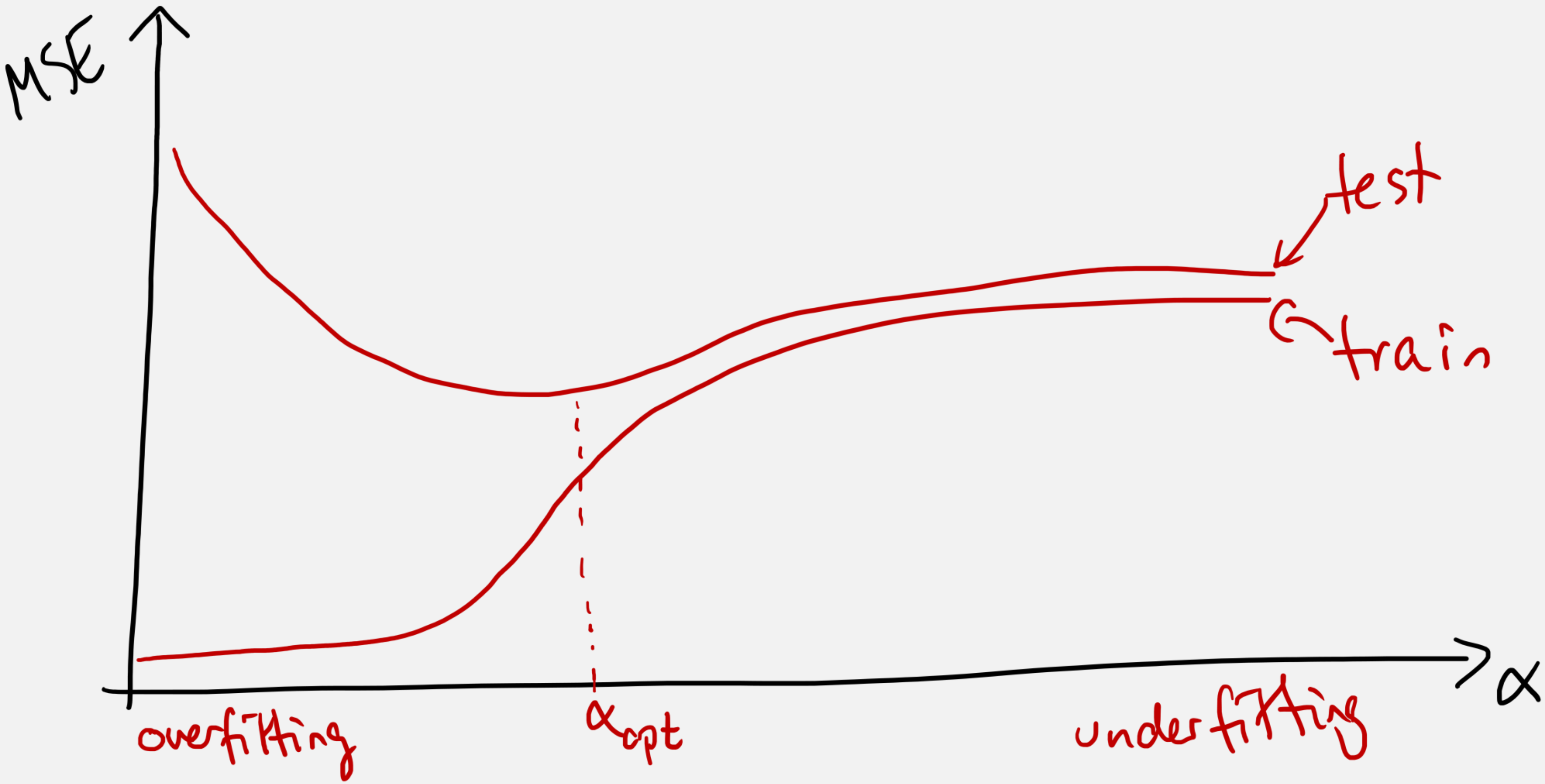
regularization parameter

penalty function of β

✓ $R(\beta) = \sum_i \beta_i^2$ ← L_2 regularization / Ridge regression

✓ $R(\beta) = \sum_i |\beta_i|$ ← L_1 regularization / Lasso regression

THE OPTIMAL REGULARIZATION PARAMETER



RIDGE VS LASSO REGRESSION

Ridge

drives coefficients to small values overall

Lasso

drives certain coefficients to zero

$\rightarrow$ built-in feature selection

Elastic Net combines the two:

$$R(\beta) \sim \gamma \cdot \text{Lasso} + (1-\gamma) \cdot \text{Ridge}$$

CODE EXAMPLE

Jupyter Notebook **Regression - Hitters**

- Explain what regression is, including OLS, Ridge, Lasso and Elastic Net regression
- Calculate and interpret relevant performance metrics
- Solve regression problems with sklearn