

## Generalization

① KNN  $\rightarrow$  Number of  $k$



$k$  kecil  $\rightarrow k = 1, 3 \rightarrow$  overfitting

$k$  lebih besar  $\rightarrow k = 7, 9, \dots \rightarrow$  mengurangi kecenderungan overfitting.

Complexity	Bias	Variance
$\uparrow$	$\downarrow$	$\uparrow$
$\downarrow$	$\uparrow$	$\downarrow$

② Decision tree  $\rightarrow$  control the model's complexity

- max - depth
- min - samples - split / leaf
- min - impurity - decrease

### ③ Linear models $\rightarrow$ Linreg, logreg

- Number of features
- Regularization  $\rightarrow$  mengatur weight / nilai coe f.
- Polynomial  $\rightarrow$  mengubah model linear menjadi non-linear



order .



overfitting

$$B_1 x_1 \rightarrow x_1 x_2, x_1^2, x_2^2$$

Regularization → mengatasi overfitting pada linear models.

$$\hat{y} = \beta_0 + \underset{\downarrow}{\beta_1^{\lambda}} x_1 + \underset{\downarrow}{\beta_2^{\lambda}} x_2 \dots \beta_n x_n$$

$\lambda \rightarrow$  penalty rate

- ① Ridge → membuat nilai coef mendekati 0 jika fitur dianggap tidak signifikan terhadap label.
- ② Lasso → membuat nilai coef menjadi 0 untuk fitur yang dianggap tidak signifikan.  
→ melakukan feature selection.
- ③ Elastic Net

# ① Ridge Regression $\rightarrow$ $L_2$ regularization

$$\propto \sum_{i=1}^n \theta_i^2$$

$\rightarrow$  menambahkan yang di dalam kotak ini ke dalam cost function model linear.

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \propto \sum_{i=1}^n \theta_i^2$$

$\xrightarrow{\text{starts from } \beta_1} \beta_0 \text{ tidak terkena regularization.}$

$$= \text{MSE}(\theta) + \propto \sum_{i=1}^n \theta_i^2$$

- Hyperparameter  $\propto$  mengontrol seberapa besar regularisasi yang akan diterapkan pada model.

$\propto = 0 \rightarrow$  Linear regression

$\propto$  sangat besar  $\rightarrow$  menjadikan nilai coef semakin mendekati 0.

- Wajib melakukan scaling pada data sebelum melakukan regularization.

e.g. MPG  $\begin{cases} \text{cylinder} \leftrightarrow 2, 4, 6, 8 \dots \rightarrow \text{skala satuan} \\ \text{mileage} \rightarrow \text{ratusan, ribuan, } \underline{\text{ratus ribuan}}. \end{cases}$

$$\hat{y} = \beta_0 - \beta_1 x_1 + \beta_2 x_2$$

$\sim 0.2 \cdot x_1 \sim 0.14 \cdot x_2$   
 $\downarrow$  hasilnya kecil       $\downarrow$  hasilnya bisa jutaan...

## ② Lasso Regression ( $L_1$ regularization)

Loss function  
Lasso  $\rightarrow J(\theta) = \text{MSE}(\theta) + \alpha \sum_{i=1}^n |\theta_i|$

←  
dikontrol juga  
oleh  $\alpha$

→ bisa menjadikan nilai coef menjadi 0  
- Feature selection

- Semakin tinggi  $\alpha$ , bisa juga semakin banyak fitur yang dihilangkan.
- Semakin kecil nilai coef, semakin tidak penting fitur tersebut.

### ③ Elastic Net

→ Lasso

→ Ridge

$$J(\theta) = \text{MSE}(\theta) + \underbrace{r}_{\checkmark} \propto \sum_{i=1}^n |\theta_i| + \frac{1-r}{2} \propto \sum_{i=1}^n \theta_i^2$$

$r$  = mix ratio

$r = 0 \rightarrow$  ridge

$r = 1 \rightarrow$  Lasso

$r = 0.5 ? \rightarrow$  balanced?

① Linreg  $\times$

② Ridge  $\checkmark$

③ Lasso

④ Elastic net



- Ridge is a good default.

- Kalau mendapati fitur yang sebenarnya tidak signifikan, kita bisa menggunakan Lasso / Elastic Net