Fifth lecture

Agenda

- Bias
- Variance
- Bias and variance trade off
- Overfitting
- Underfitting
- Regularization (lasso and ridge regression)

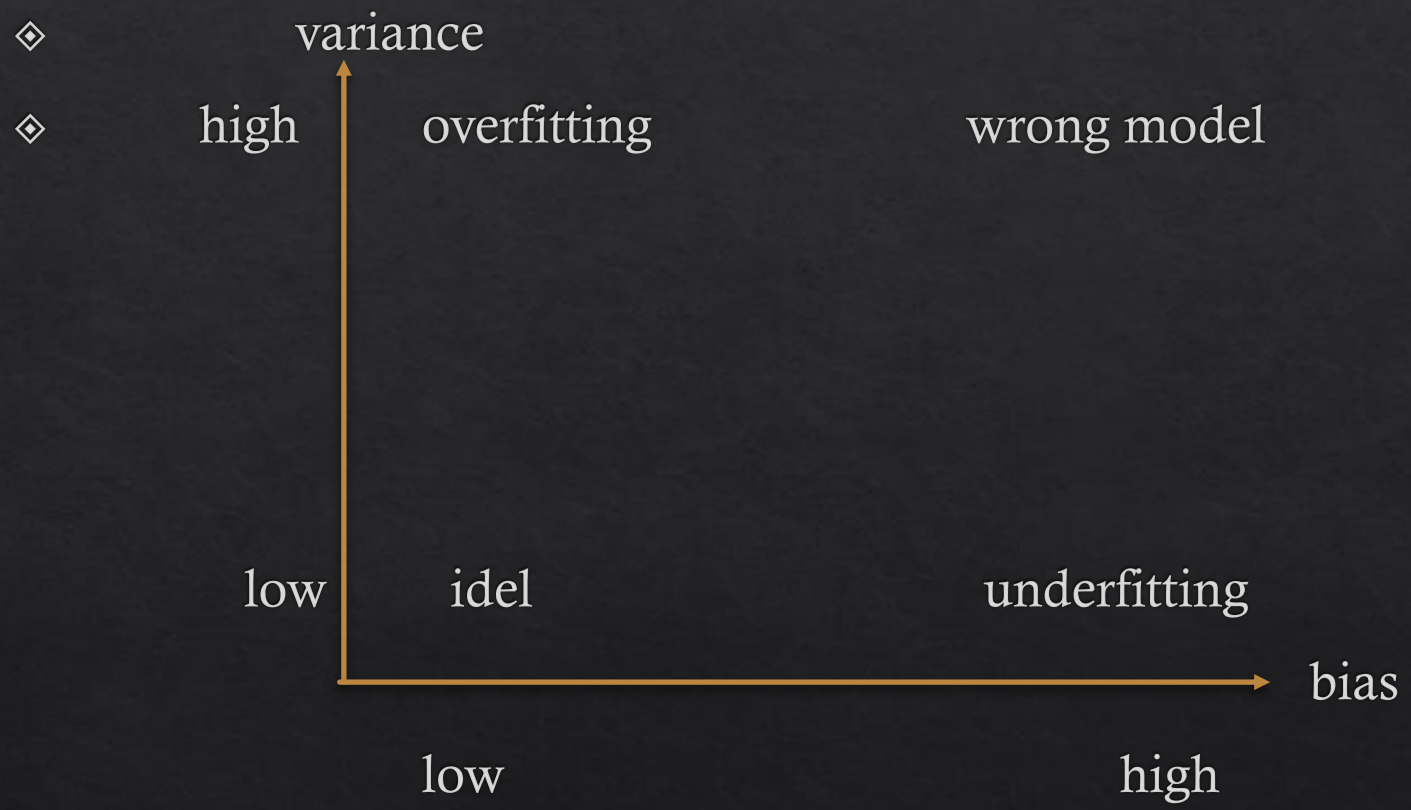
Bias and Variance

model will have good accuracy when it is trying to make predictions on new or unseen data
for example, using the data which is not included in the training set.
Good accuracy also means that the value predicted by the model will be very much close to the actual value.
Bias will be low and variance will be high when model performs well on the training data but performs bad or poorly on the test data.

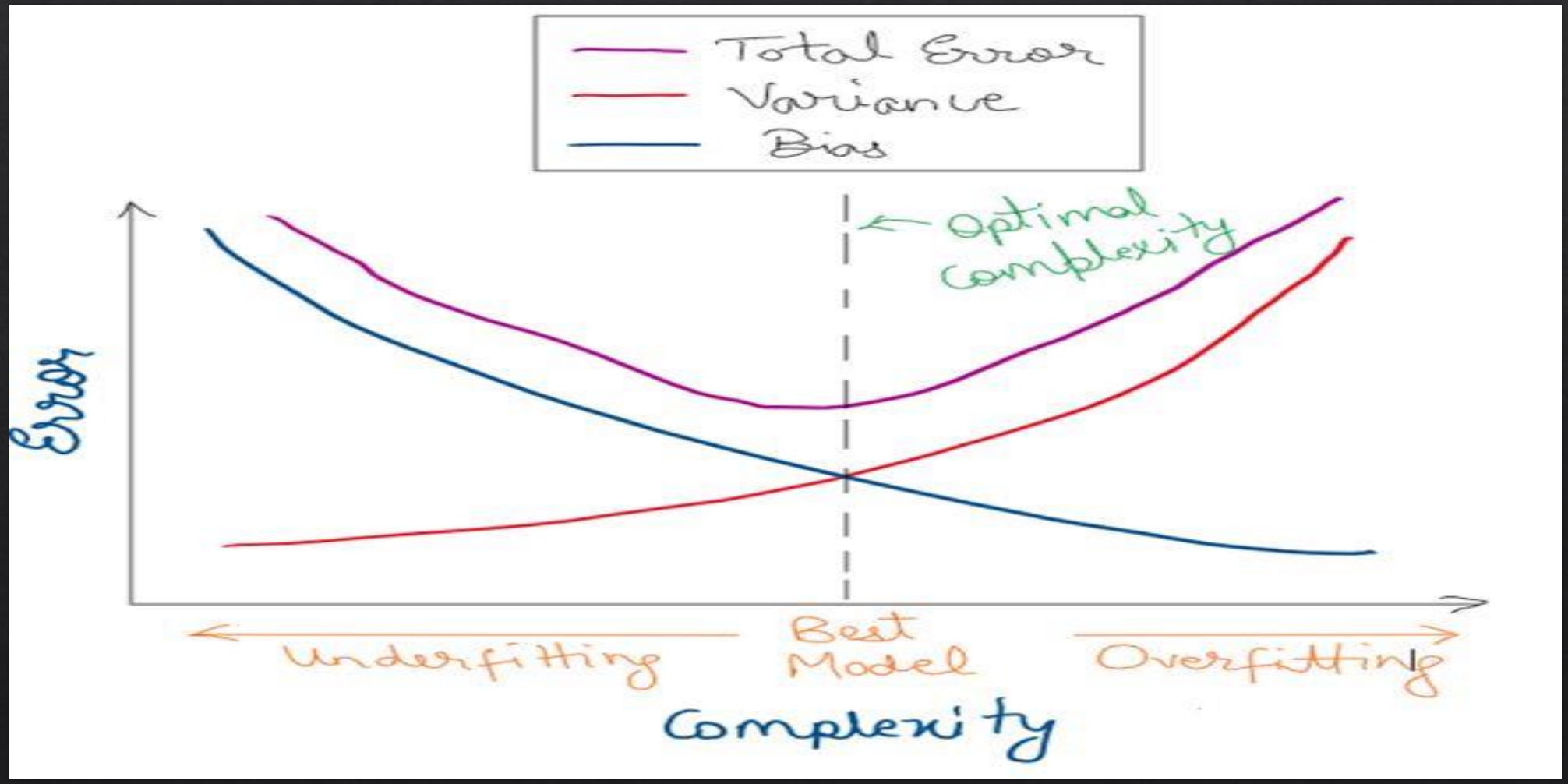
High variance means the model cannot generalize to new or unseen data. (This is the case of overfiting)
If the model performs poorly (means less accurate and cannot generalize) on both training data and test data,
it means it has high bias and high variance (This is the case of underfiting)
If model performs well on both test and training data.
Performs well meaning, predictions are close to actual values for unseens data so accuracy will be high.
In this case, bias will be low and variance will also be low.
The best model must have low bias (low error rate on training data) and low variance (can generalize and has low error rate on new or test data) (This is the case for best fit model)

so always have low bias and low variance for your models.

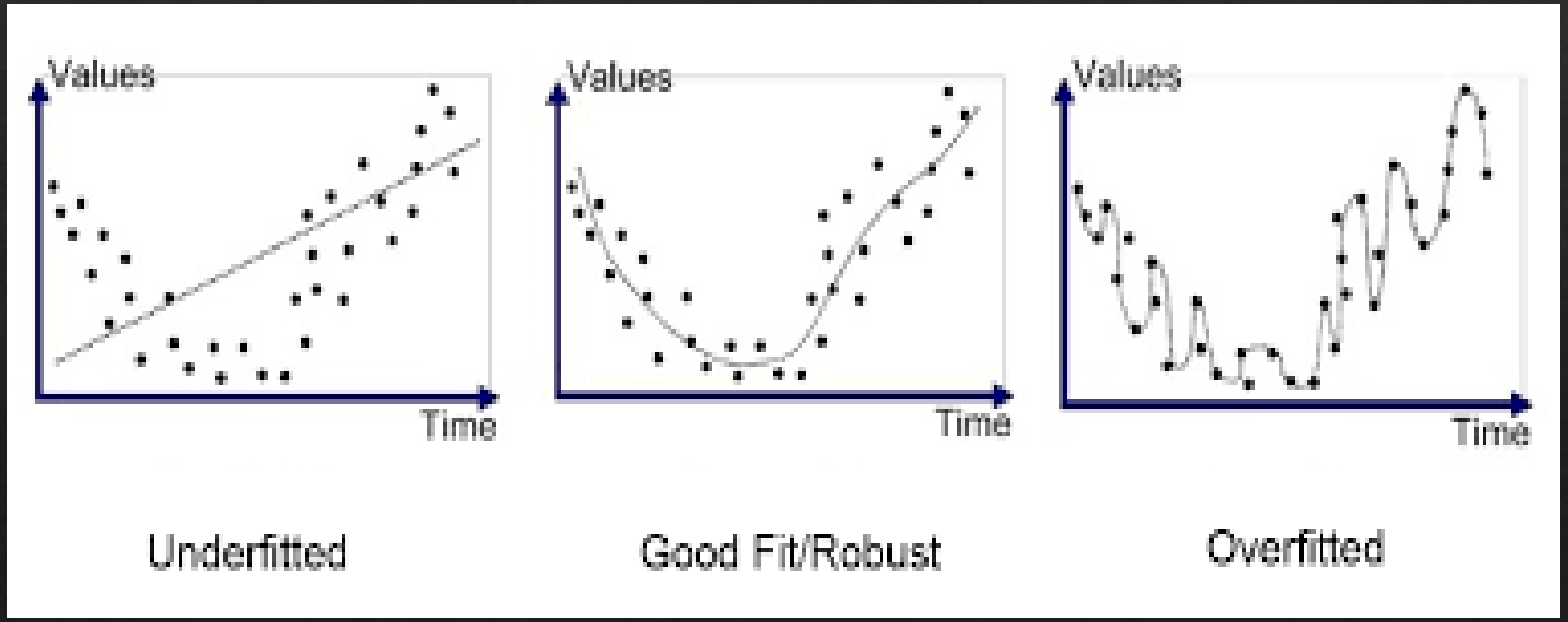
Bias and Variance



Bias and variance trade off



There are many method to git rid of overfitting one of them is called Regularization (lasso L1 and ridge L2 regression)



Lasso and ridge regression equation

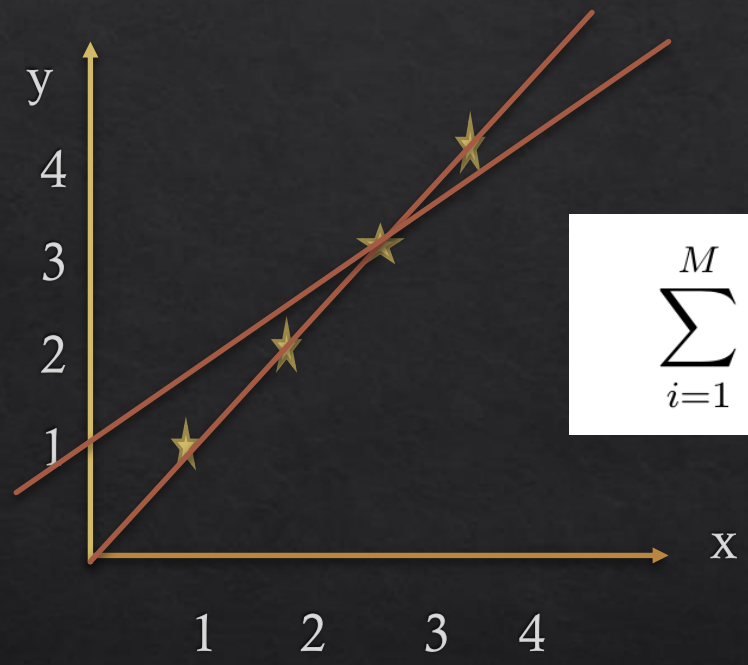
- **ridge regression formula**

$$\sum_{i=1}^{M}(y_i-\hat{y}_i)^2=\sum_{i=1}^{M}\left(y_i-\sum_{j=0}^{p}w_j\times x_{ij}\right)^2+\lambda\sum_{j=0}^{p}w_j^2 \tag{1.3}$$

- **lasso regression formula**

$$\sum_{i=1}^{M}(y_i-\hat{y}_i)^2=\sum_{i=1}^{M}\left(y_i-\sum_{j=0}^{p}w_j\times x_{ij}\right)^2+\lambda\sum_{j=0}^{p}|w_j| \tag{1.4}$$

Example
let's rid of overfitting using Ridge



let line equation y = 2x → the overfit line

second line y = 1.6 * x +1

$$\sum_{i=1}^{M} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{M} \left(y_i - \sum_{j=0}^{p} w_j \times x_{ij} \right)^2 + \lambda \sum_{j=0}^{p} w_j^2 \qquad (1.3)$$

lambada is a value between 0 and 1

for the overfit line = 0+1*2^2 = 4

The shifted line

x	y	(y-y_pred)	(y-y_pred)^2
1	2	2.6	0.36
2	4	4.2	0.04
3	6	5.8	0.04
4	8	7.4	0.36
Sum of square difference			0.8

- Penalty = 1 * 1.6^2 = 2.56

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 + \lambda \sum_{j=0}^{p} w_j^2 \qquad (1.3)$$

- 0.8 + 2.56 = 3.56
-
- so we success to rid of the overfit we can reduce again by changing the lambada value