universität freiburg

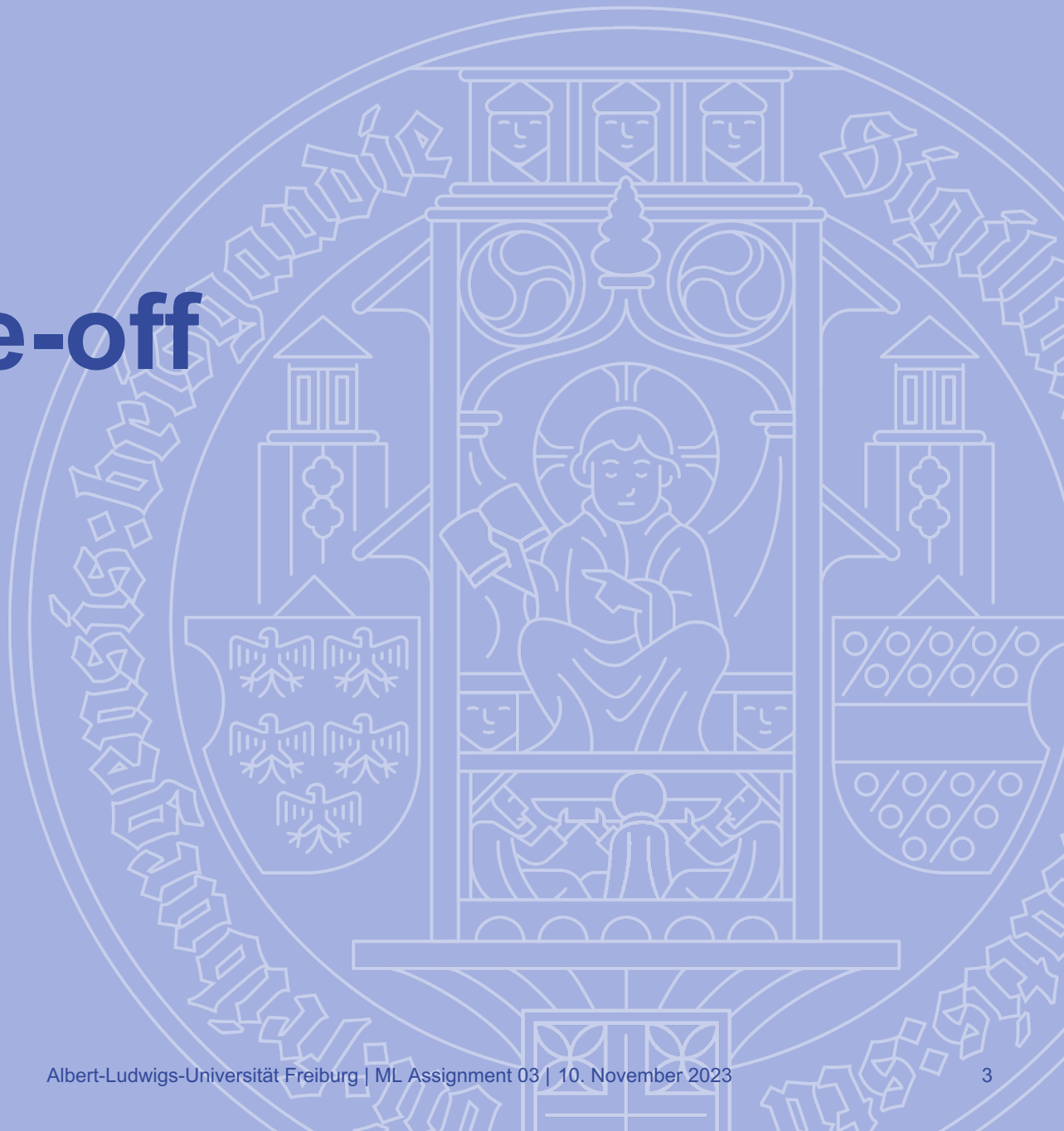
Assignment 03

Machine Learning WS23/24

Assignment 03
Solution

1. Bias-Variance Trade-off
2. Overfitting and Underfitting in Linear Regression

Bias-Variance Trade-off



Bias-Variance Trade-off

$$E_{x,y,D}\left[\left(\hat{f}(x;D)-y\right)^2\right] =$$

$$\underbrace{E_{x,D}\left[\left(\hat{f}(x;D)-\bar{f}(x)\right)^2\right]}_{\text{Variance}} + \underbrace{E_{x,y}\left[\left(\bar{f}(x)-\bar{y}(x)\right)^2\right]}_{\text{Bias}^2} + \underbrace{E_{x,y}\left[\left(\bar{y}(x)-y\right)^2\right]}_{\text{Noise}}$$

where $\hat{f}(\cdot)$ is the prediction model,

$\bar{f}(\cdot)$ is the expected prediction model,

y is the target,

and $\bar{y}(x)$ is the expected target.

Bias-Variance Trade-off

$$E_{x,y,D}\left[\left(\hat{f}(x;D)-y\right)^2\right] =$$

$$\underbrace{E_{x,D}\left[\left(\hat{f}(x;D)-\bar{f}(x)\right)^2\right]}_{\text{Variance}} + \underbrace{E_{x,y}\left[\left(\bar{f}(x)-\bar{y}(x)\right)^2\right]}_{\text{Bias}^2} + \underbrace{E_{x,y}\left[\left(\bar{y}(x)-y\right)^2\right]}_{\text{Noise}}$$

Variance – measures how much a model's predictions **vary** across different training datasets. High variance suggests the model heavily depends on its training data and may not perform well on new, unseen data. This is often seen in **complex**, nonlinear models which, while fitting training data well, can struggle with new data due to **overfitting,** i.e. interpolating the data and not learning the underlying relationship in the data. Cross-validation can help measure this by showing the model's consistency across various training subsets.

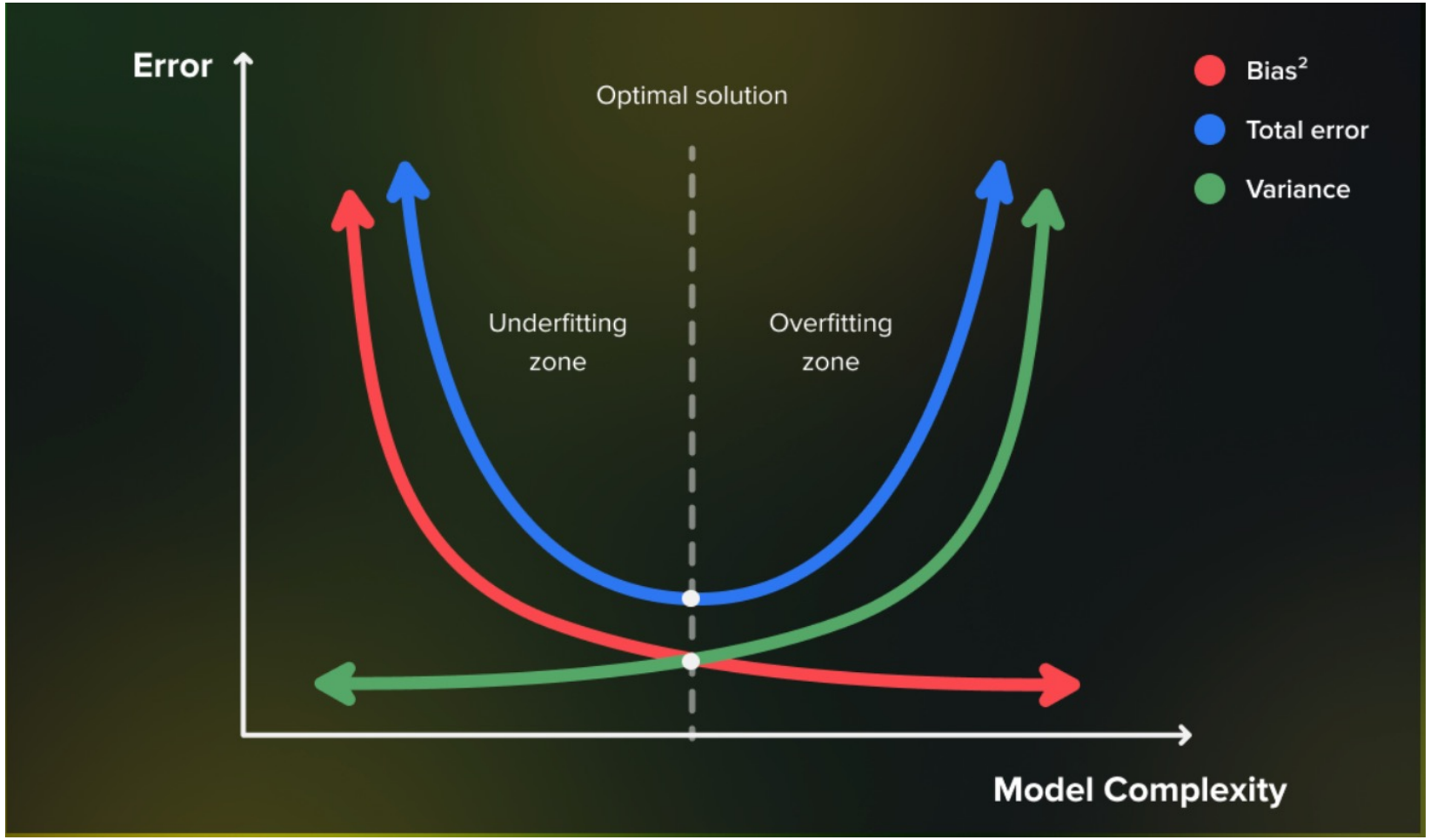
Bias-Variance Trade-off

$$E_{x,y,D}\left[\left(\hat{f}(x;D) - y\right)^2\right] =$$

$$\underbrace{E_{x,D}\left[\left(\hat{f}(x;D) - \bar{f}(x)\right)^2\right]}_{\text{Variance}} + \underbrace{E_{x,y}\left[\left(\bar{f}(x) - \bar{y}(x)\right)^2\right]}_{\text{Bias}^2} + \underbrace{E_{x,y}\left[\left(\bar{y}(x) - y\right)^2\right]}_{\text{Noise}}$$

Bias – measures how a model's predictions differ from the actual distribution of the value it tries to predict. High-bias models, which tend to oversimplify, result in noticeable errors during training and testing, indicative of **underfitting** — the model's failure to capture the true patterns and variations in the data.

Bias-Variance Trade-off



Source: https://serokell.io/blog/bias-variance-tradeoff

Bias-Variance Trade-off

How does regularization prevent overfitting?

Regularization is a technique that penalizes more complex models to prevent overfitting. By adding a constraint to the model parameters, typically by shrinking them towards zero, regularization reduces model complexity. This discourages the model from fitting the noise in the training data, leading to improved generalization on unseen data.

Bias-Variance Trade-off

Let us say now you have 500 MRI images and you have tested 4 different machine learning models whose error metrics are as follows:

	Model 1	Model 2	Model 3	Model 4
Train error [%]	0.1	25	15	0.2
Test error [%]	13	19	40	3

Bias-Variance Trade-off

	Model 1	Model 2	Model 3	Model 4
Train error [%]	0.1	25	15	0.2
Test error [%]	13	19	40	3
Bias	**Low**	**High**	**High**	**Low**

Bias-Variance Trade-off

	Model 1	Model 2	Model 3	Model 4
Train error [%]	0.1	25	15	0.2
Test error [%]	13	19	40	3
Bias	Low	High	High	Low
Variance	**High**	**High**	**High**	**Low**

Bias-Variance Trade-off

	Model 1	Model 2	Model 3	Model 4
Train error [%]	0.1	25	15	0.2
Test error [%]	13	19	40	3
Bias	Low	High	High	Low
Variance	High	High	High	Low
Fitting	**Overfitting**	**Underfitting**	**Underfitting**	**Good fit**

Bias-Variance Trade-off

In the case of model 2, what could be the reason that the test error is lower than the training error?

	Model 2
Train error [%]	25
Test error [%]	19

Bias-Variance Trade-off

In the case of model 2, what could be the reason that the test error is lower than the training error?

	Model 2
Train error [%]	25
Test error [%]	19

- There could be complex instances to learn in the training set.
- The samples in the test dataset could be from a different source than the training dataset and have easier instances.