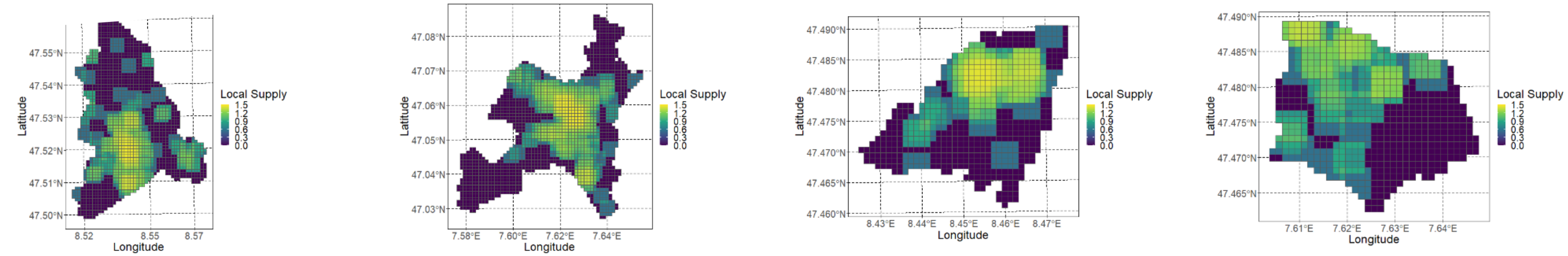
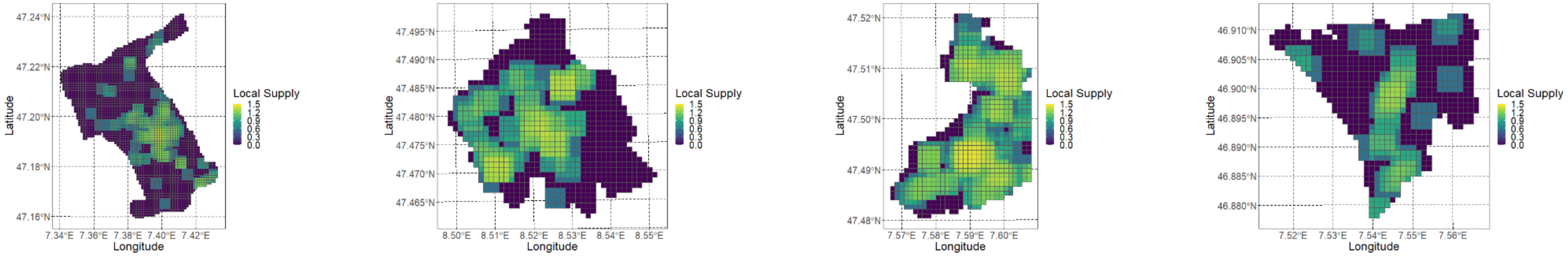
Fachhochschule Graubünden
University of Applied Sciences

Regularization and Optimization

Dr. Yves Staudt



(a) Grenchen | (b) Oberglatt | (c) Reinach | (d) Rubigen



(e) Buelach | (f) Burgdorf | (g) Dielsdorf | (h) Dornach

Lernziel

Die Studierende sind in der Lage

- Deep Learning Modelle gegen Overfitting vorzubeugen.

FH GR

Regularization

- Regularization is a technique used to **prevent** overfitting
- **Overfitting** occurs when a model is too complex and fits the training data too closely, resulting in poor performance on new, unseen data
- Regularization improves the **generalization performance** of a model on unseen data
- Regularization adds a penalty term to the **loss function** of a neural network in order to prevent overfitting
- Regularization helps to prevent the model from fitting the training data **too closely**
- Regularization improves its ability to **generalize** to new data
- The model is encouraged to **generalize better** to new data
- The **penalty term** in regularization encourages the model to use smaller weights

L1 and L2 regularization

- These methods add a **penalty term** to the loss function of the deep learning model
- **L1 regularization** adds the absolute value of the weights to the loss function
- **L2 regularization** adds the squared weights
- This encourages the model to **use smaller weights**, which can help prevent overfitting
- The strength of the regularization is controlled by a hyperparameter called the **regularization parameter**
- The regularization parameter determines how much **weight** is given to the penalty term

L1 Regularization

Mathematically, the loss function with L1 regularization can be written as:

$$Loss = Data\ Loss + \lambda \cdot |w|$$

where

- $Data\ Loss$ is the usual loss function used to train the neural network,
- w is the vector of weights,
- $|w|$ is the L1 norm of the weights, and
- λ is the regularization parameter

The L1 norm is simply the sum of the absolute values of the weights.

L2 Regularization

Mathematically, the loss function with L2 regularization can be written as:

$$Loss = Data\ Loss + \lambda \cdot ||w||^2$$

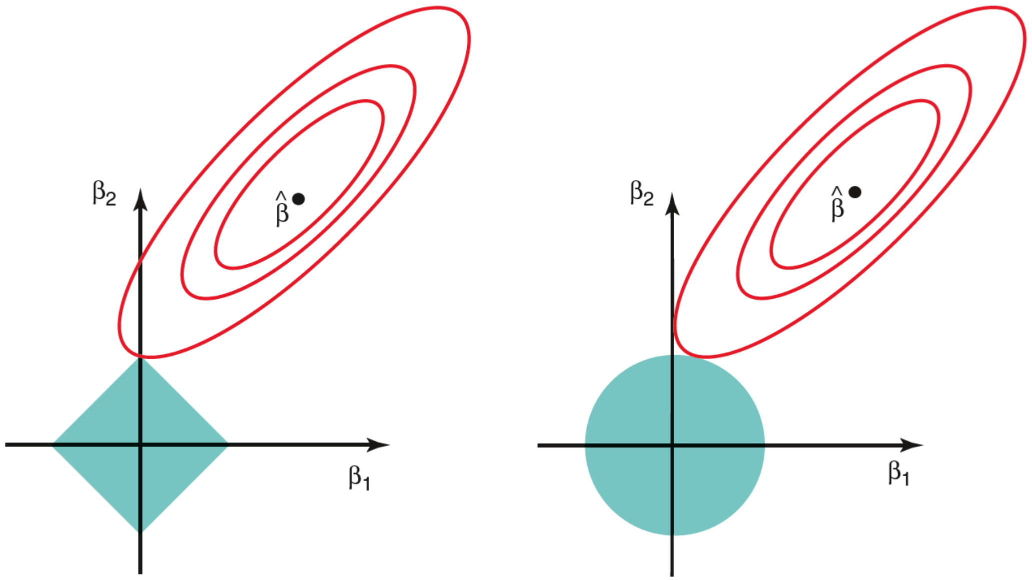
where

- $Data\ Loss$ is the usual loss function used to train the neural network,
- w is the vector of weights,
- $||w||$ is the L2 norm of the weights, and
- λ is the regularization parameter

L2 norm is the square root of the sum of the squared weights

FH
GR

L1 and L2 regularization



Visualization of the contours error and constraint functions for the lasso (left) and ridge regression (right) (James et al., 2013).
The solid blue areas are the constraint regions, $|\beta_1| + |\beta_2| \leq s$ and $\beta_1^2 + \beta_2^2 \leq s$, while the red ellipses are the contours of the RSS.

- L1 regularization tends to encourage sparse weight vectors
- In L1 regularization many of the weights will be exactly zero
- L1 regularization is useful for feature selection, as it effectively removes some of the less important features from the model
- L2 regularization tends to spread the weights more evenly across all features.

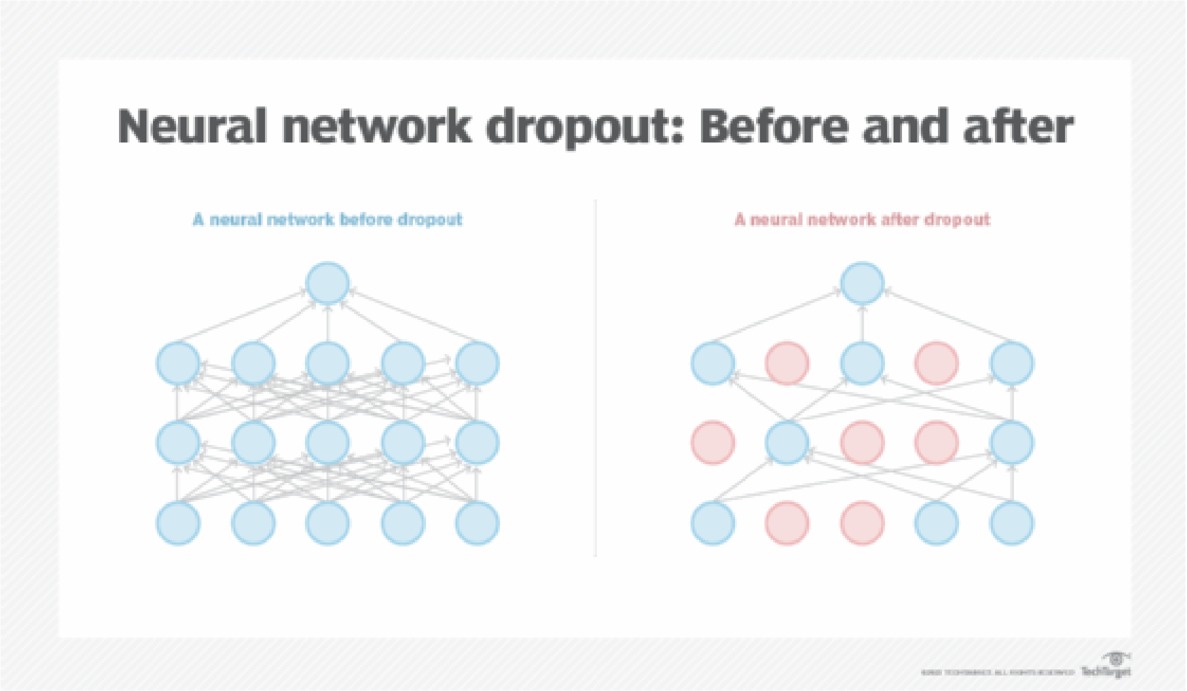
FH GR

Dropout

Dropout is a regularization method that randomly drops out a fraction of the deep learning in a layer during training

Dropout forces the remaining neurons to learn more robust features

Dropout prevents the model from relying too heavily on any one neuron

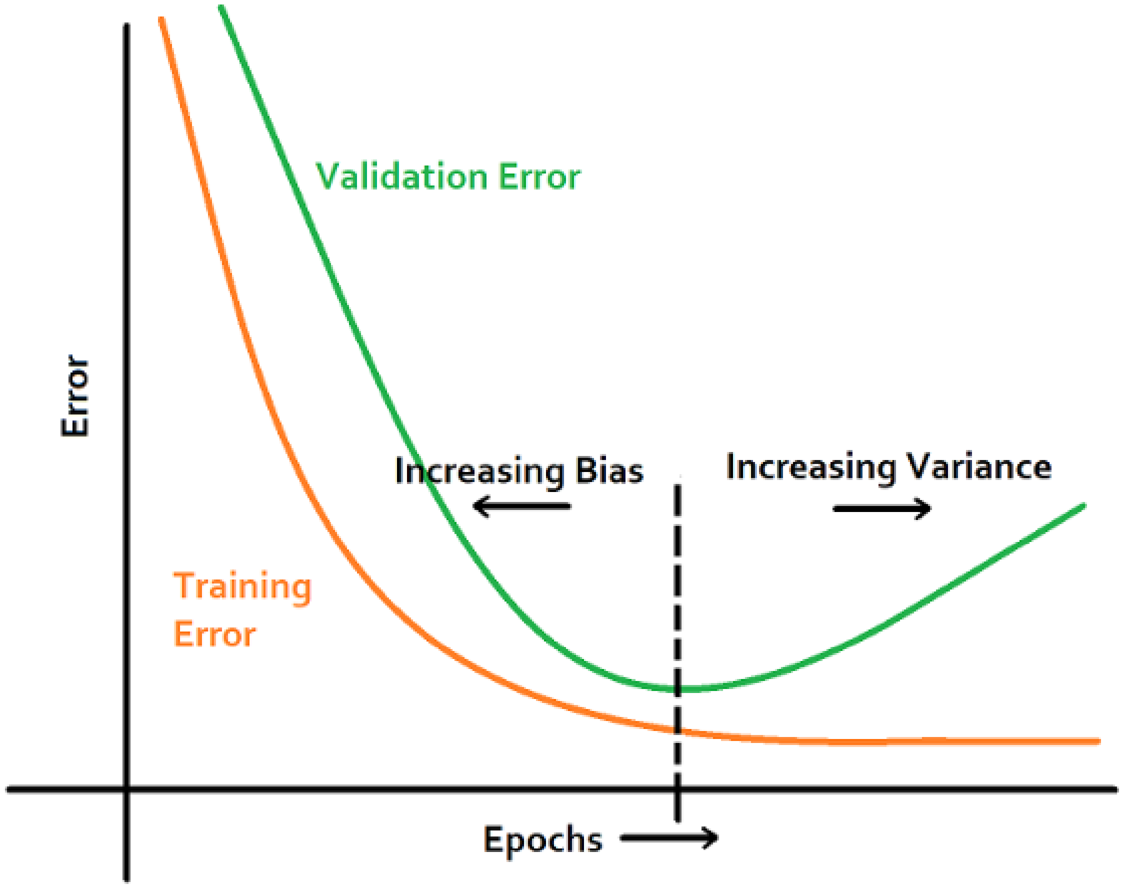


FH GR

Early Stopping

Early stopping is a technique that stops the training process when the performance of the model on a validation set stops improving

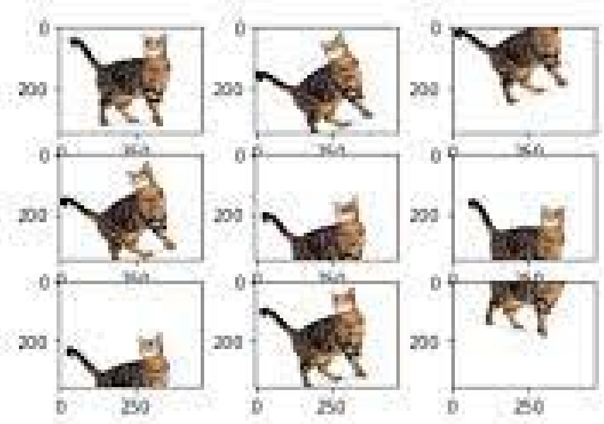
Early stopping prevents the model from continuing to overfit to the training data.



FH
GR

Data Augmentation

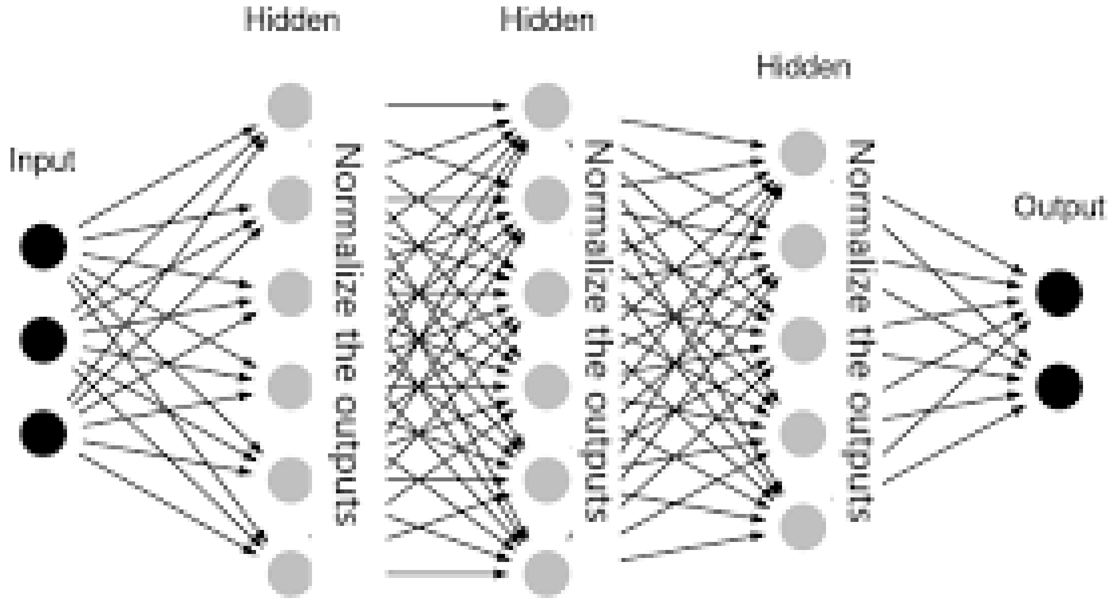
- Data augmentation is a technique that involves generating new training data by applying various transformations to the existing data.
- Data augmentation helps the model to learn more robust features
- Data augmentation improves its ability to generalize to new data



FH
GR

Batch Normalization

- Batch normalization is a technique that normalizes the inputs to a layer in order to stabilize the distribution of the activations
- Batch normalization helps to prevent overfitting
- Batch normalization improves the generalization performance of the model



FH
GR

Referenzen

- Chollet, F. (2019), Deep Learning with Keras, Manning.

Fragen

Darstellung eines Fragesymbol aufgerufen von der Webseite https://www.qnigge.de/news/detail/modul-v/#images am 12.07.2021.

Fachhochschule Graubünden
Pulvermühlestrasse 57
7000 Chur
T +41 81 286 24 24
info@fhgr.ch

Vielen Dank für Ihre Aufmerksamkeit.
Grazia fitg per l'attenziun.
Grazie per l'attenzione.