NLP and Deep Learning
MAT3399

Lecture 10:
Some Guidelines for Training Deep Learning Models

Tuan Anh Nguyen @ Aimesoft
ted.nguyen95@gmail.com

Content taken from Andrew Ng's Deep Learning Specialization Course on Coursera

Training ML/DL is a highly iterative process

Data

Numbers of layers

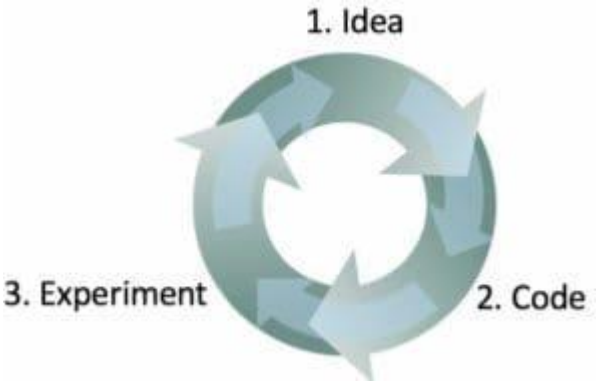
Numbers of units

Learning rate

Activation functions

Architecture

...



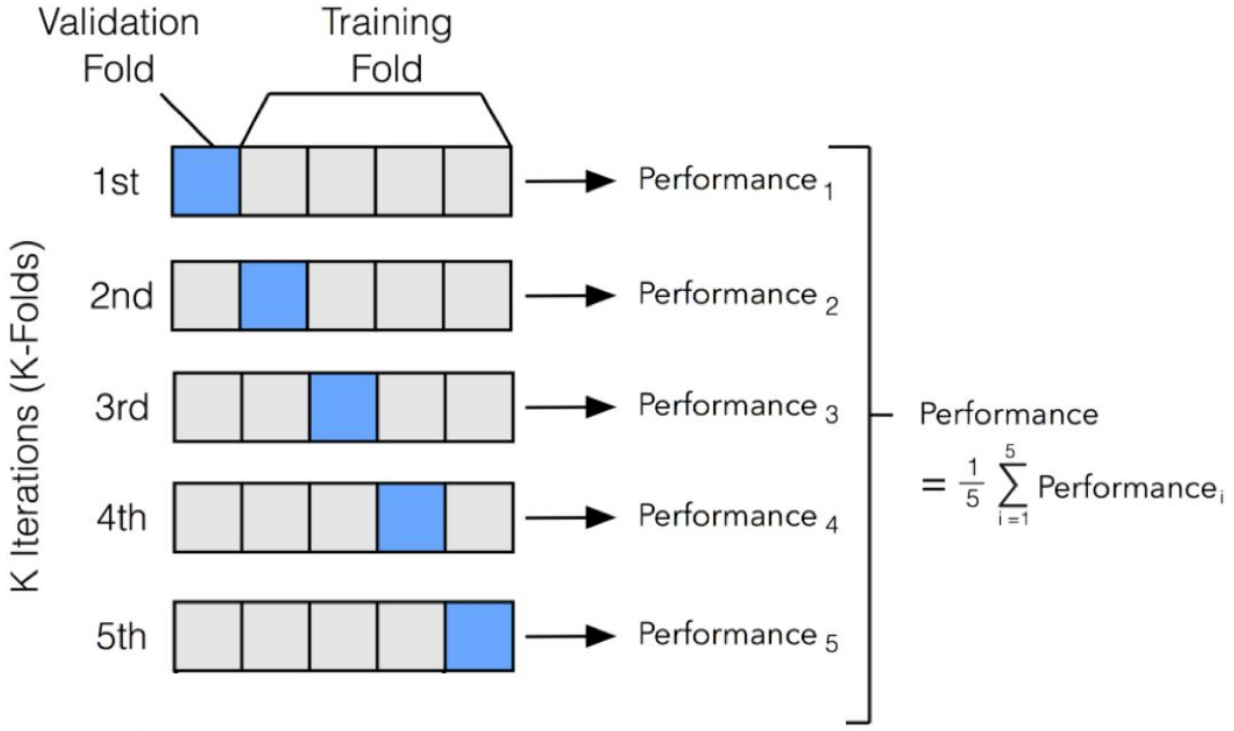
Build a model fast, then iterate

Train / Dev (Validation) / Test

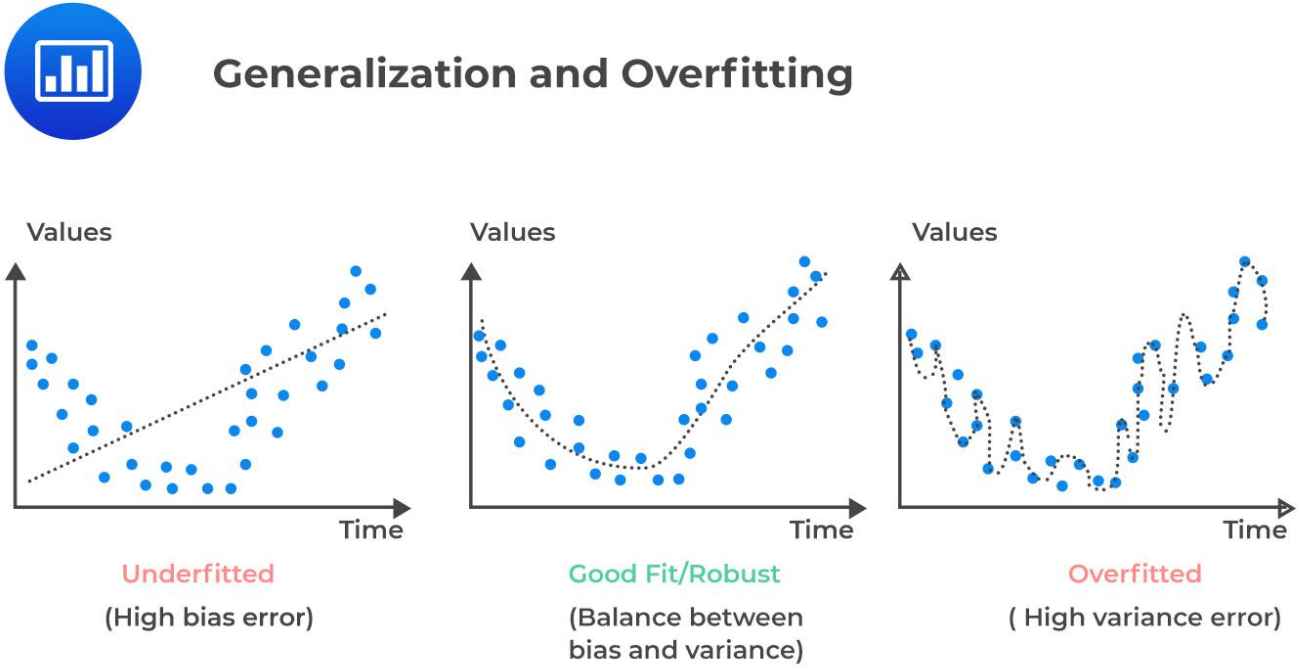
What is the best ratio for train/dev/test dataset?

Test dataset and train dataset should have same distribution

Cross Validation



Bias and Variance



Basic Recipe for Deep Learning

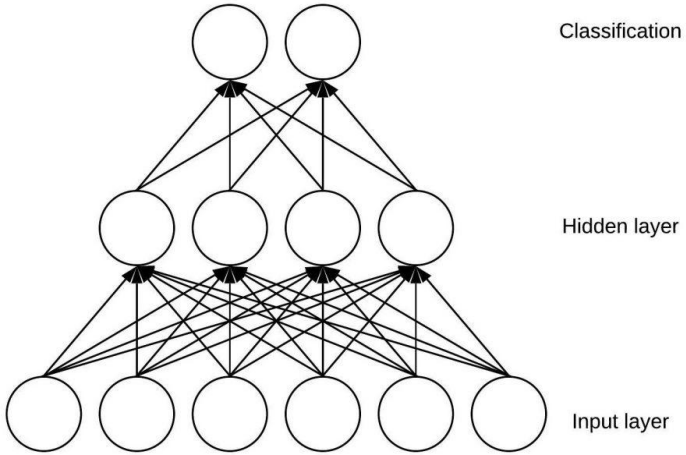
Problem	Possible solutions
High bias (Underfitting)	- Use more complex networks - Train for more epochs
High variance (Overfitting)	- Collect more data - Use regularization `

Understanding Regularization

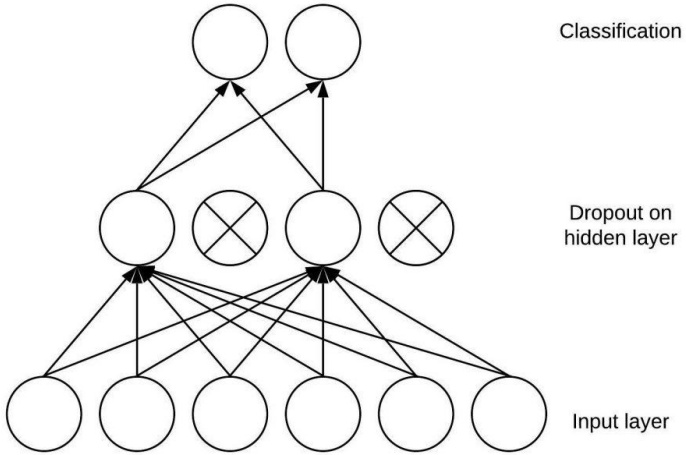
L2 Regularization

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} w_i^2$$

Dropout



Without Dropout



With Dropout

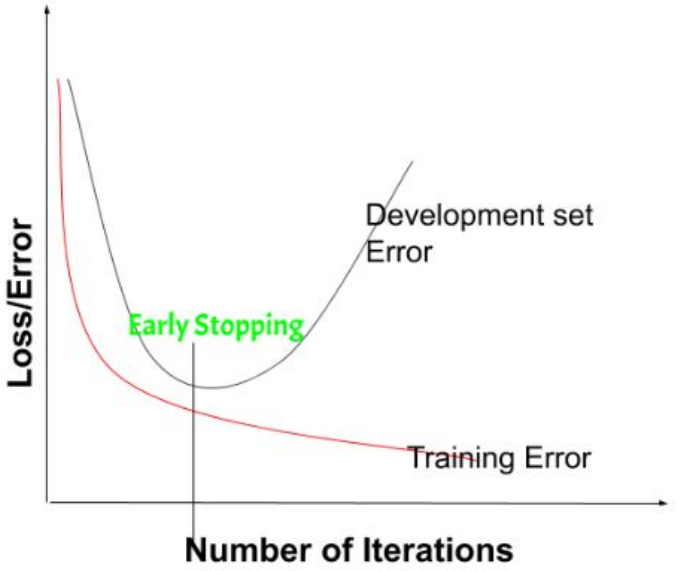
Data Augmentation

Data augmentation is a technique that increase the training dataset artificially

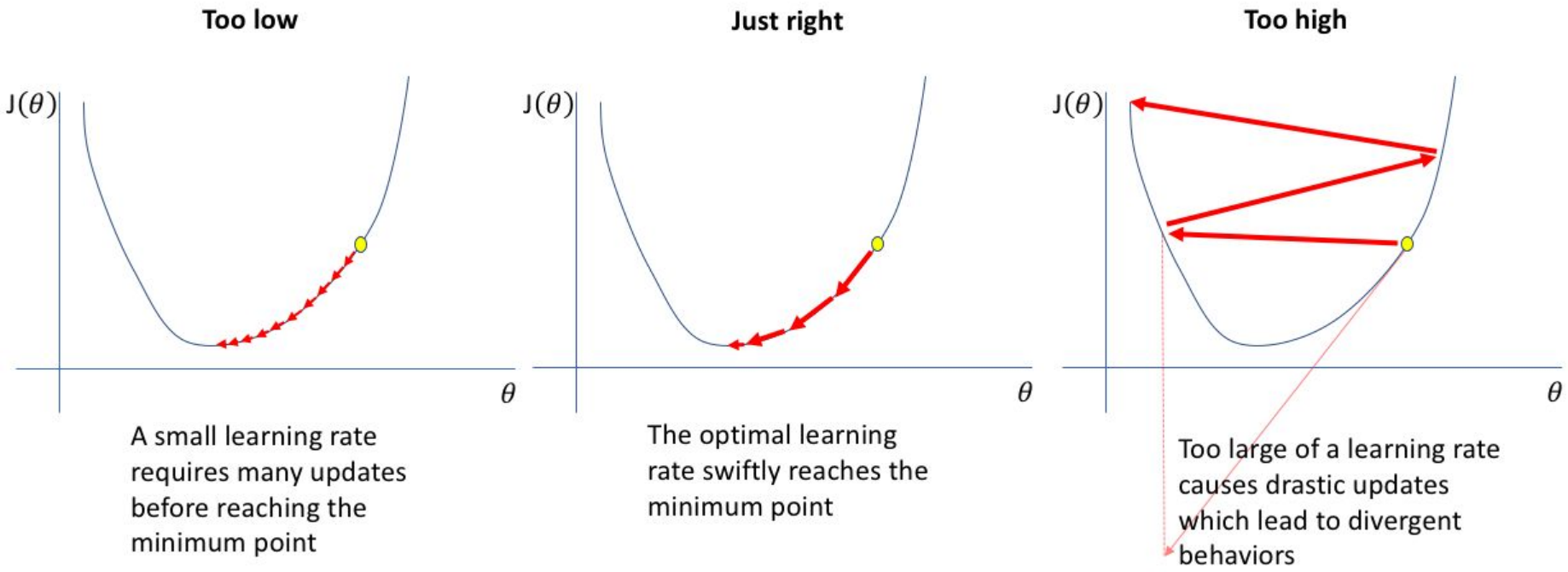
Some methods for data augmentation in NLP:

- Synonym replacement: Replace a word by its synonyms
- Contextual word augmentation: Use masked language model (like BERT) to insert a word randomly at a random position
- Back translation: Use machine translation model to translate text to another language then translate back to the original language
- Or you can just use ChatGPT

Early Stopping



Decay Learning Rate

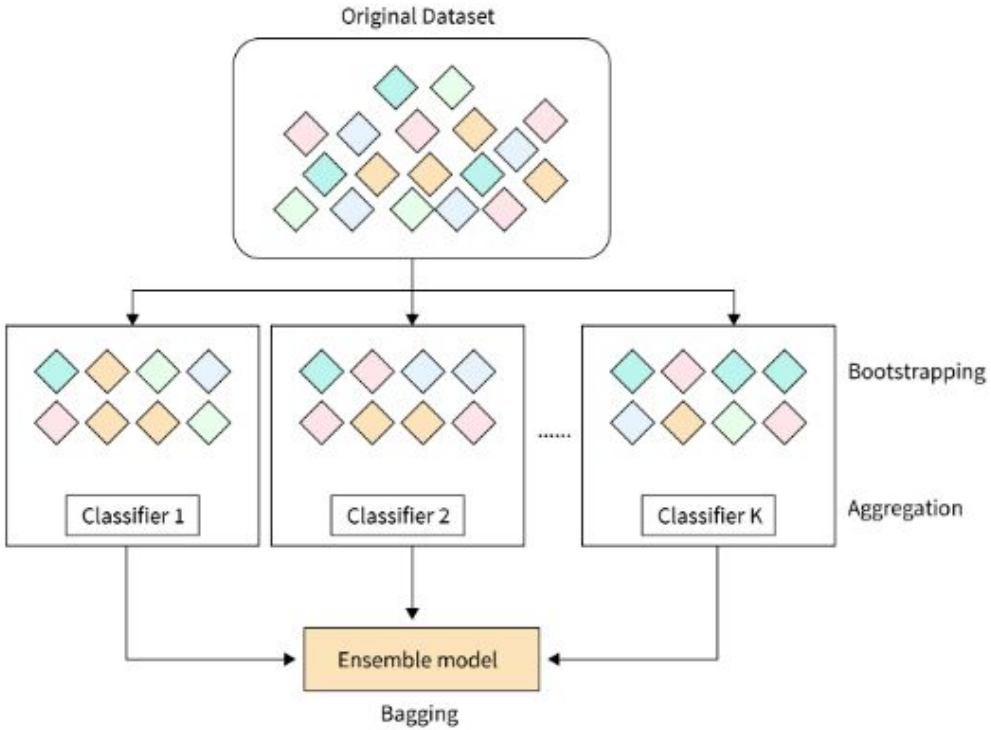


Error Analysis

It is helpful to do these tasks when you want to improve your deep learning model:

- Data quality assessment
- Confusion matrix analysis: This will help you identify if the model is struggling with specific classes or types of data.
- Error categorization: Group errors into categories based on their characteristics
- Comparative analysis: Compare your model's performance with baseline models or alternative approaches

Model Combining Methods – Model Bagging



Model Combining Methods – Ensemble Learning