Better Predictions with Ensemble Learning

Christophe Zoghbi
CEO @ Zaka

About me

Christophe Zoghbi

- Software Engineer
- Computer Vision & Machine Learning consultant
- Founder & President of **Beirut AI**
- Founder & CEO of **Zaka**

LinkedIn: christophezoghbi
Twitter: @kristoffzoghbi
Facebook: christophe.zoghbi

ZAKA
Educate ▸ Educate ▸ Educate
An Intelligent Transformation

Outline

- Problem definition
- Case for Ensemble Learning
- What is Ensemble Learning?
- How to Ensemble Neural Network Models?
  - Different techniques
  - Different options
- Hands-on code examples

https://github.com/zaka-ai/wonderlandai-workshop

ZAKA
Educate ▸ Educate ▸ Educate
An Intelligent Transformation

High Variance of Neural Network Models

- Training deep neural networks can be **very computationally expensive**!
- After the investment of so much **time and resources**, there is **no guarantee** that the final model will have low generalization error and perform well on examples not seen during training.
- Neural Networks are sensitive to initial conditions:
  - Initial random weights
  - Statistical noise in the training dataset
- This means that each time a model is trained, it may learn a slightly (**or dramatically**) different version of the mapping function from inputs to outputs, and have different performance on the training and holdout datasets

Notebook

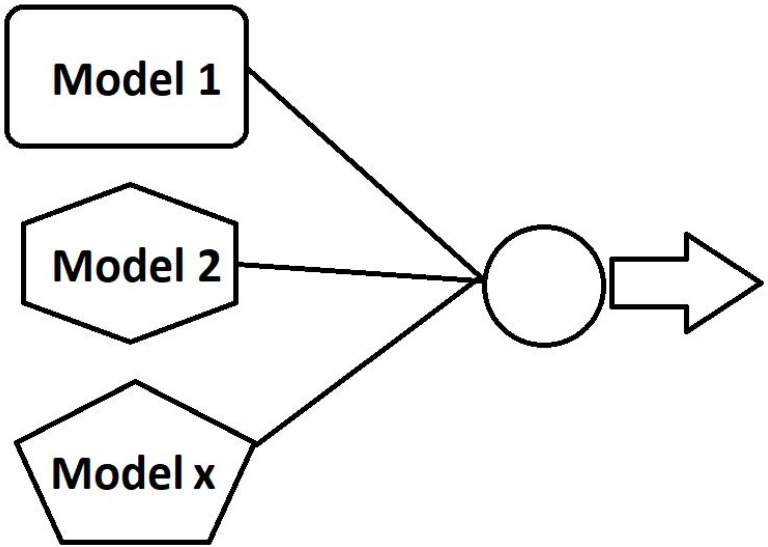
00_MLP_Variance

ZAKA

Educate ▸ Educate ▸ Educate

AN INTELLIGENT TRANSFORMATION

Reduce Variance Using an Ensemble of Models



A solution to the high variance of neural networks is to **train multiple models** and **combine their predictions**.

Why Ensemble Learning?

- Model averaging works because different models **will usually not** make all the same errors on the test set.
- The ensemble can also result in **better predictions** than any single best model!
- This approach is used by winners in machine learning competitions (Kaggle)

How to Ensemble Neural Network Models?

- A committee of networks (what?)
- Ensembles may be as small as 3,5, or 10 trained models
- There are many variations:
  - **Training Data**: Vary the choice of data used to train each model in the ensemble.
  - **Ensemble Models**: Vary the choice of the models used in the ensemble.
  - **Combinations**: Vary the choice of the way that outcomes from ensemble members are combined.

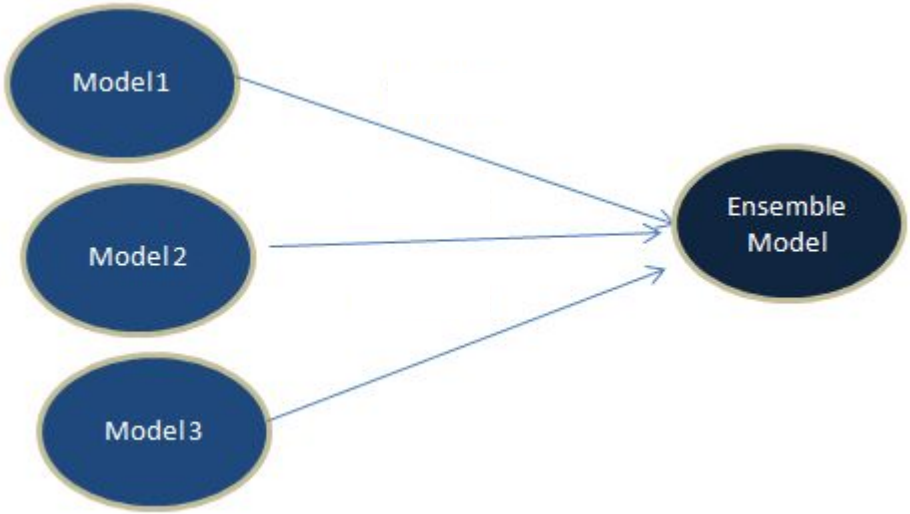
Model Averaging Ensemble

1. Train multiple models
2. Combine predictions

   a. Regression task

   b. Binary classification

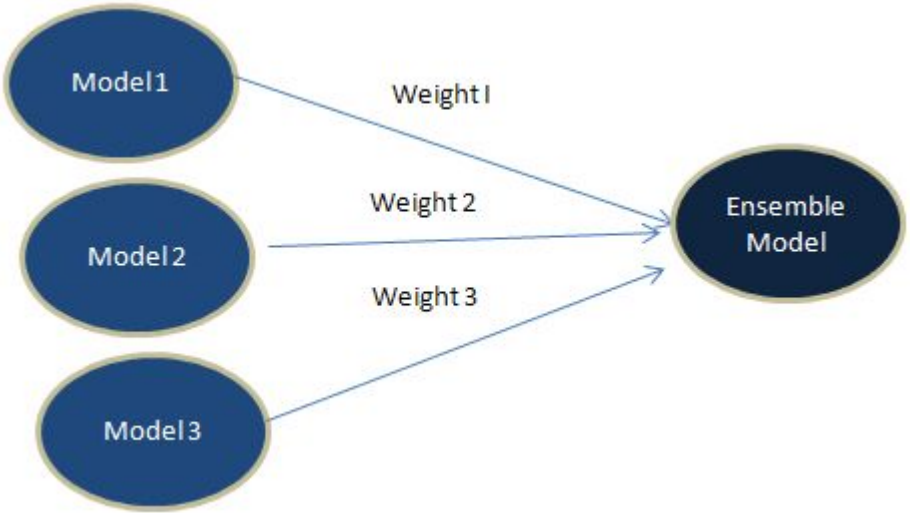
   c. Multiclass classification



Notebook

01_Model_Averaging_Ensemble

Weighted Average Ensemble



How do we find the weights?

- Grid search weight values between 0 and 1 for each ensemble member
- Use a directed optimization process!

Notebook

02_Weighted_Average_Ensemble

Resampling Ensembles

- **Random Splits**. The dataset is repeatedly sampled with a random split of the data into train and test sets.
- **k-fold Cross-Validation**. The dataset is split into k equally sized folds, k models are trained and each fold is given an opportunity to be used as the holdout set where the model is trained on all remaining folds.
- **Bootstrap Aggregation.** Random samples are collected with replacement and examples not included in a given sample are used as the test set.

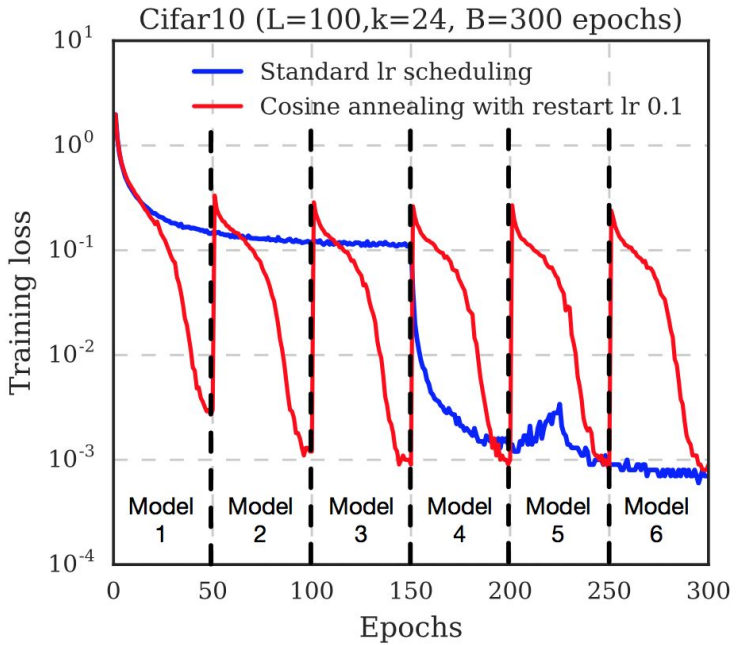
Perhaps the most widely used resampling ensemble method is bootstrap aggregation, more commonly referred to as bagging.

Notebook

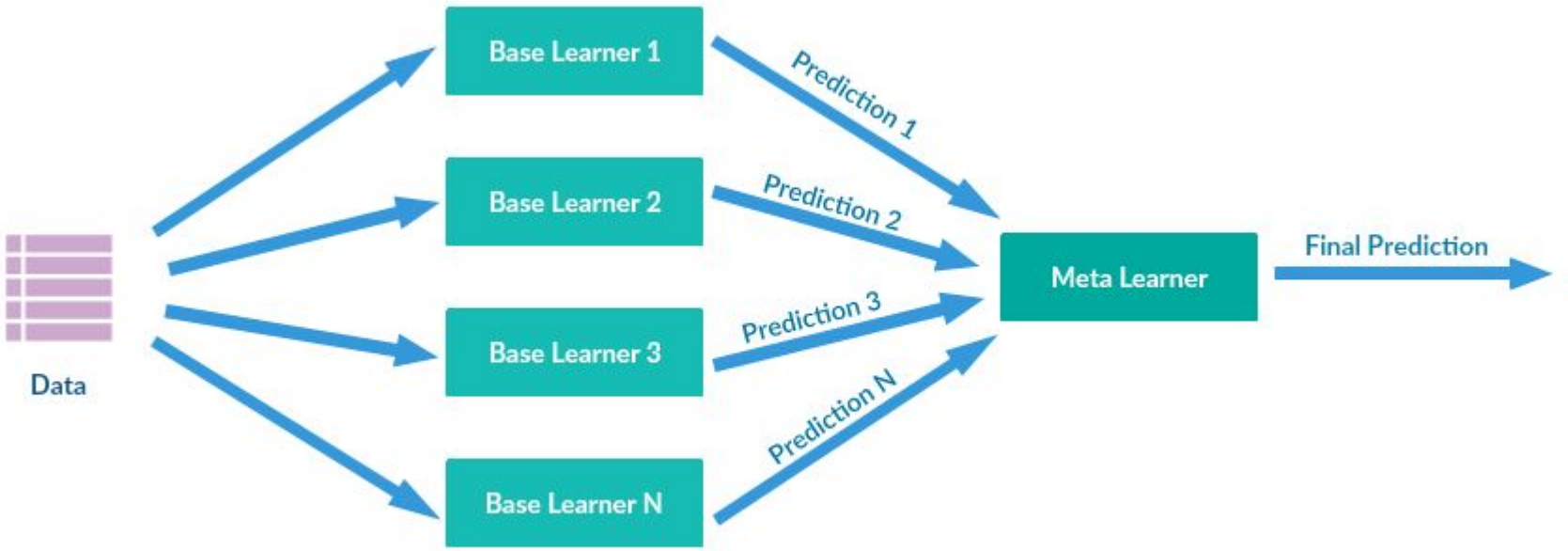
03_Resampling_Ensembles

Snapshot Ensembling

- Training Deep Neural Networks is **EXPENSIVE**!
- Collect multiple models from a single training run.
- **Downside**: the models may be good, but too similar.
- **Solution**: systematically and aggressively change the learning rate during training to result in very different network weights



Stacked Generalization Ensemble (Stacking)



Machine Learning Mastery

https://machinelearningmastery.com

Thank you!

Questions?