W207– Applied Machine Learning

Cornelia Ilin, PhD

School of Information
UC Berkeley

KNN, Decision Trees, Ensembles

Announcements

- Final project: **baseline presentation** next week during the live session. No more than 12 minutes

Baseline presentation. Your slides should include:

- Title, Authors
- What is the question you will be working on? Why is it interesting?
- What is the data you will be using? Include data source, size of dataset, main features to be used. Please also include summary statistics of your data.
- What prediction algorithms do you plan to use? Please describe them in detail.
- How will you evaluate your results? Please describe your chosen performance metrices and/or statistical tests in detail.

Announcements

- Final project: **baseline presentation** next week during the live session. No more than 12 minutes

- ~~Assignment 7~~: focus on your final project

Last week

- General concepts: FF Neural Networks
- Training, validation, and test datasets
- Application: Detect Diabetic Retinopathy using image data

Last week

- General concepts: FF Neural Networks
- Training, validation, and test datasets
- Application: Detect Diabetic Retinopathy using image data

add nonlinearities to NN

Also important:

activation functions (Relu, Tanh, Logistic, etc)

Last week

- General concepts: FF Neural Networks
- Training, validation, and test datasets
- Application: Detect Diabetic Retinopathy using image data

add nonlinearities to NN

Also important:

activation functions (Relu, Tanh, Logistic, etc)

regularization (Dropout, L1, L2, etc)

helps with overfitting

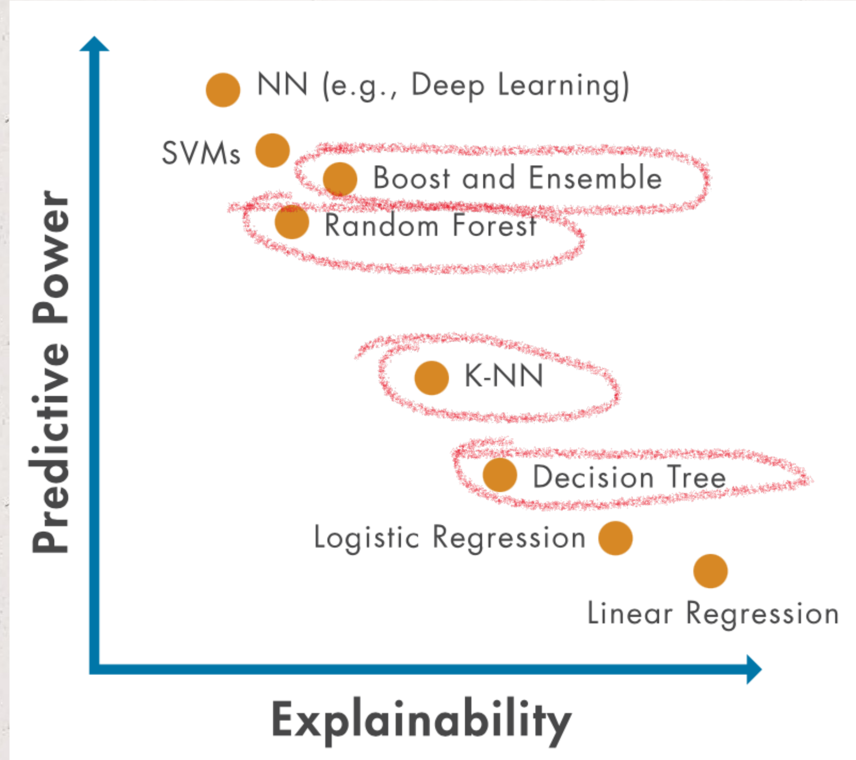
This week - breakout rooms

- As a team (write your answers in a Word document):

  - discuss pro and cons of KNN and Decision Trees
  - define and explain when Ensembles are useful
  - rank the following models based on explainability:
    * linear regression
    * logistic regression
    * FFNN
    * KNN
    * Decision trees

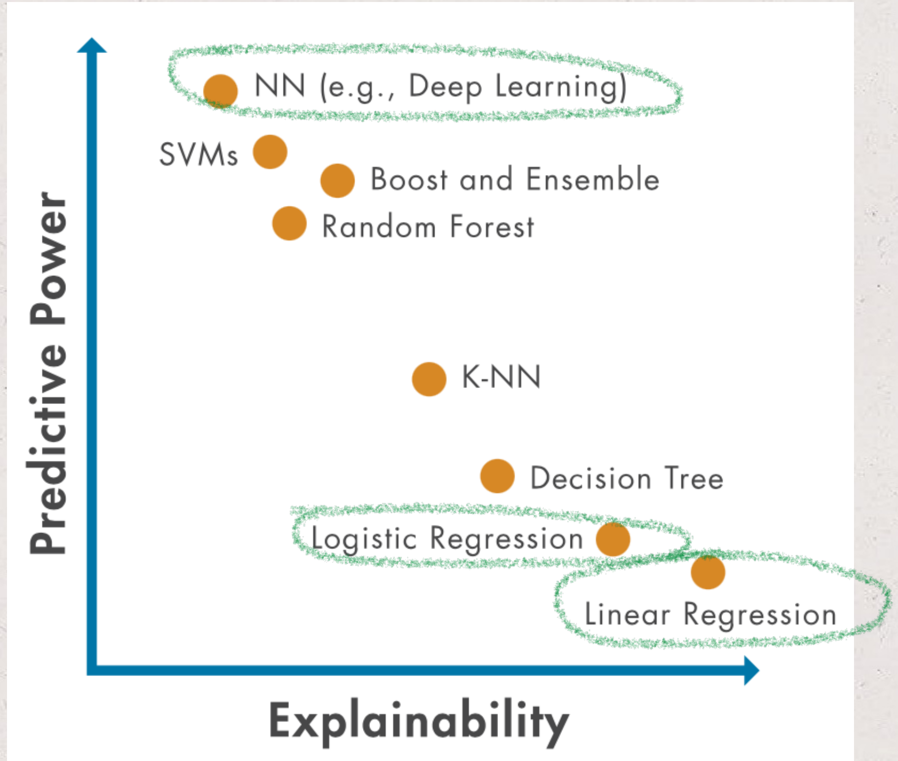
Today's learning objectives

- KNN, Decision Trees, Ensembles



v.s.

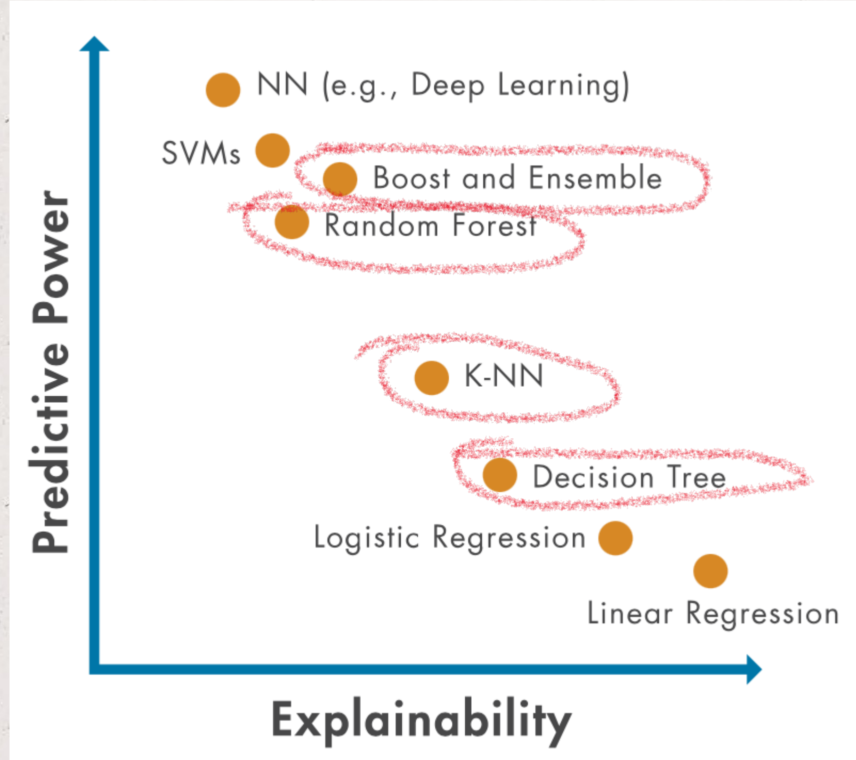
New focus of the class



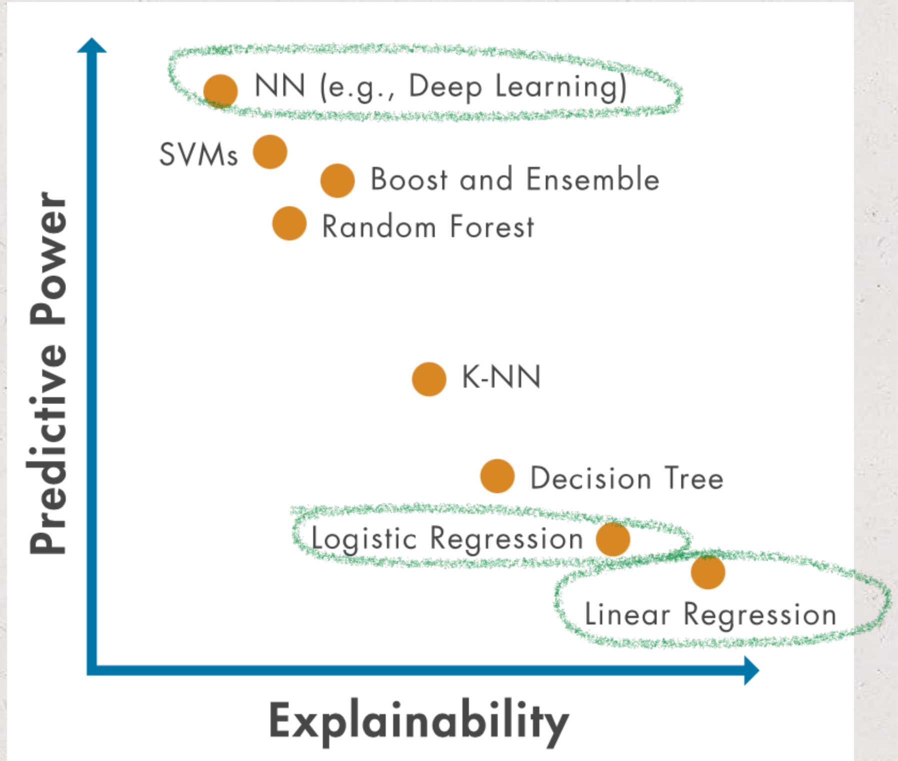
Today's learning objectives

- KNN, Decision Trees, Ensembles

New focus of the class



v.s.

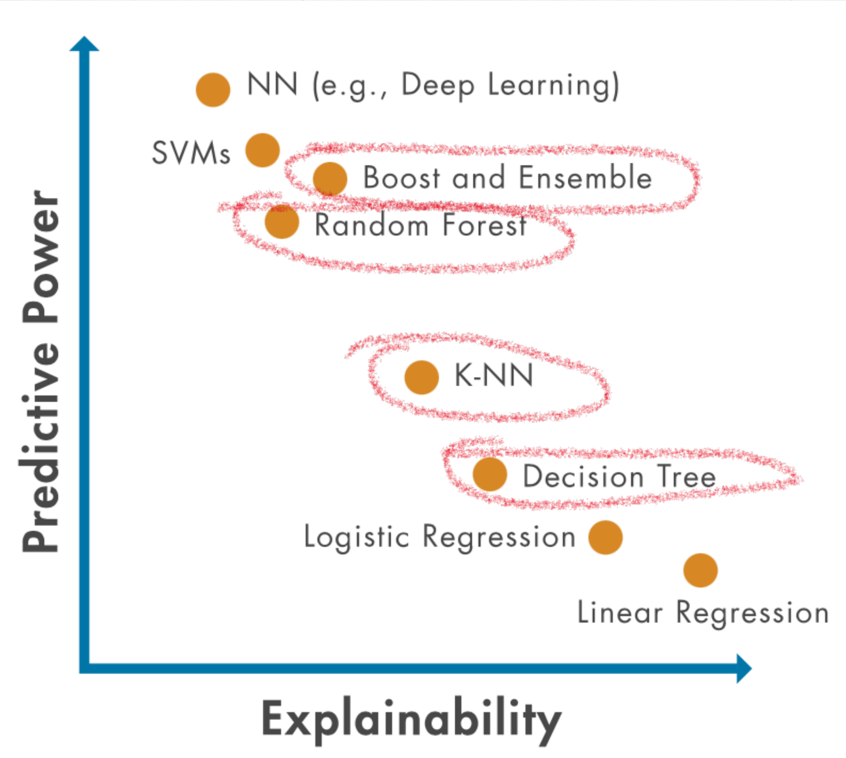


Trade-off between model performance and explainability

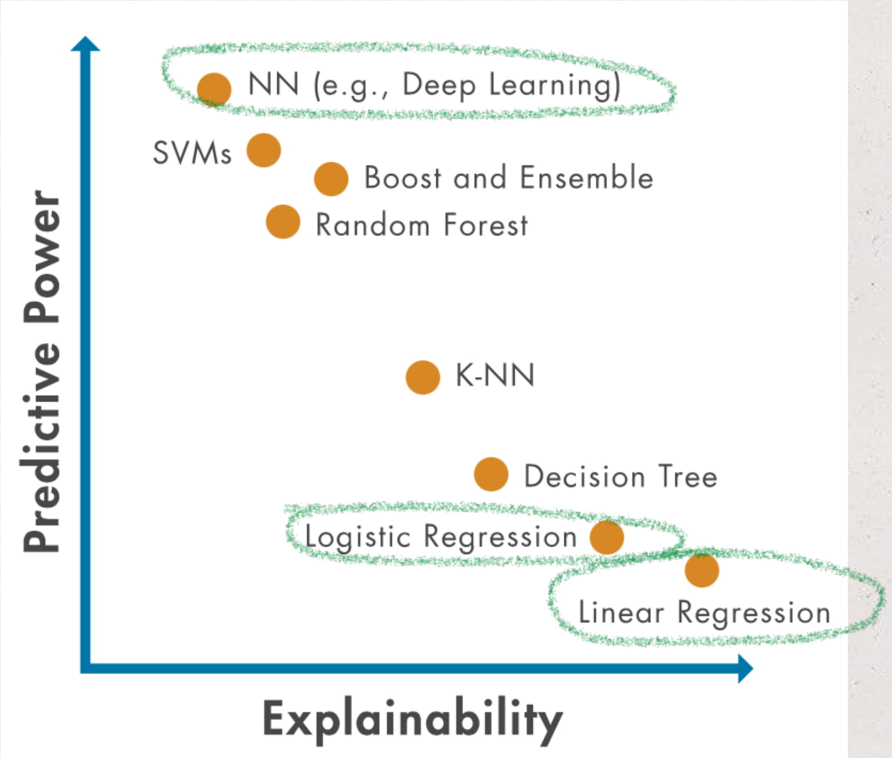
Today's learning objectives

- KNN, Decision Trees, Ensembles

New focus of the class



v.s.

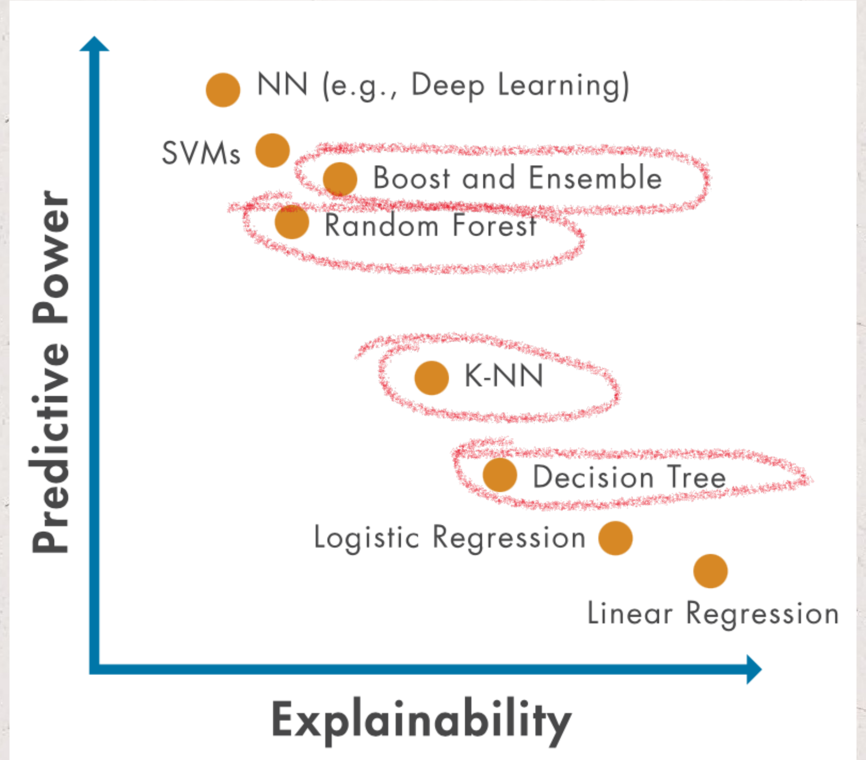


see week06

Trade-off between model performance and explainability → **Explainable AI/ML using Shap**

Today’s learning objectives

- KNN, Decision Trees, Ensembles

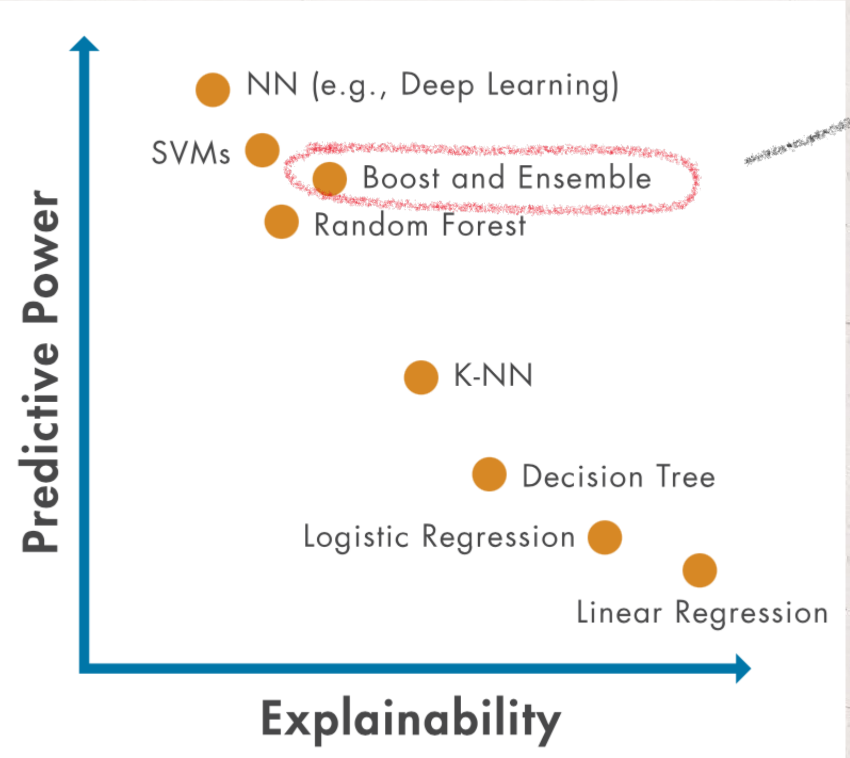


Check week07 GitHub repo

What are the main differences between these models? (other than explainability and predictive power)

Today's learning objectives

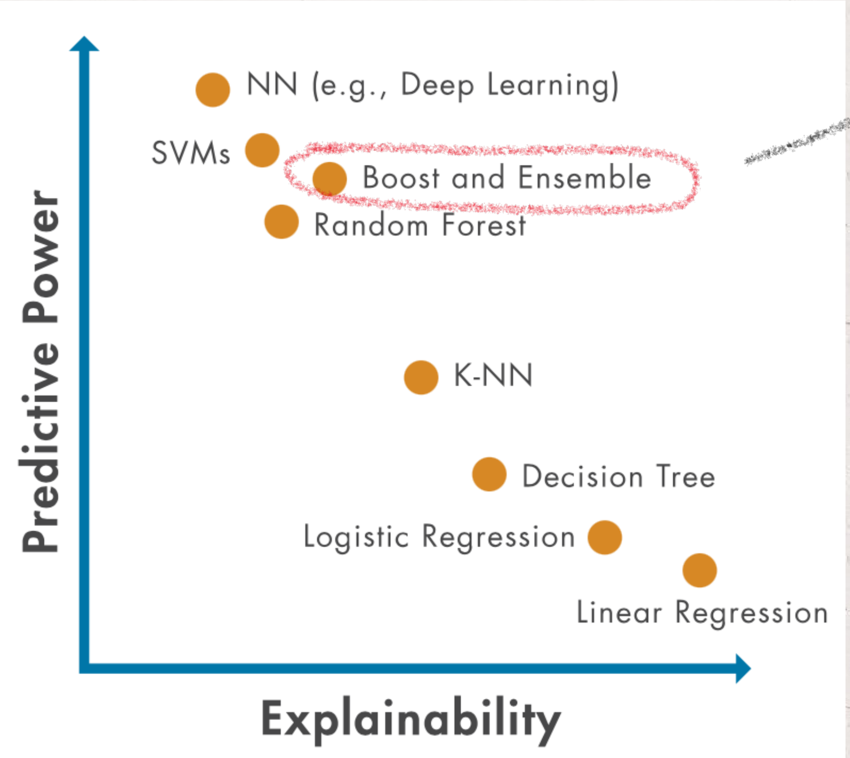
- KNN, Decision Trees, Ensembles



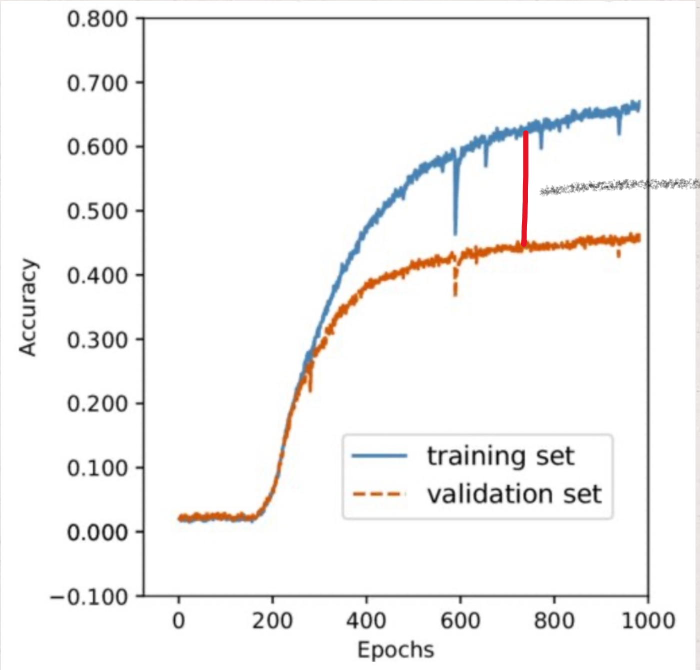
Motivation?

Today's learning objectives

- KNN, Decision Trees, Ensembles



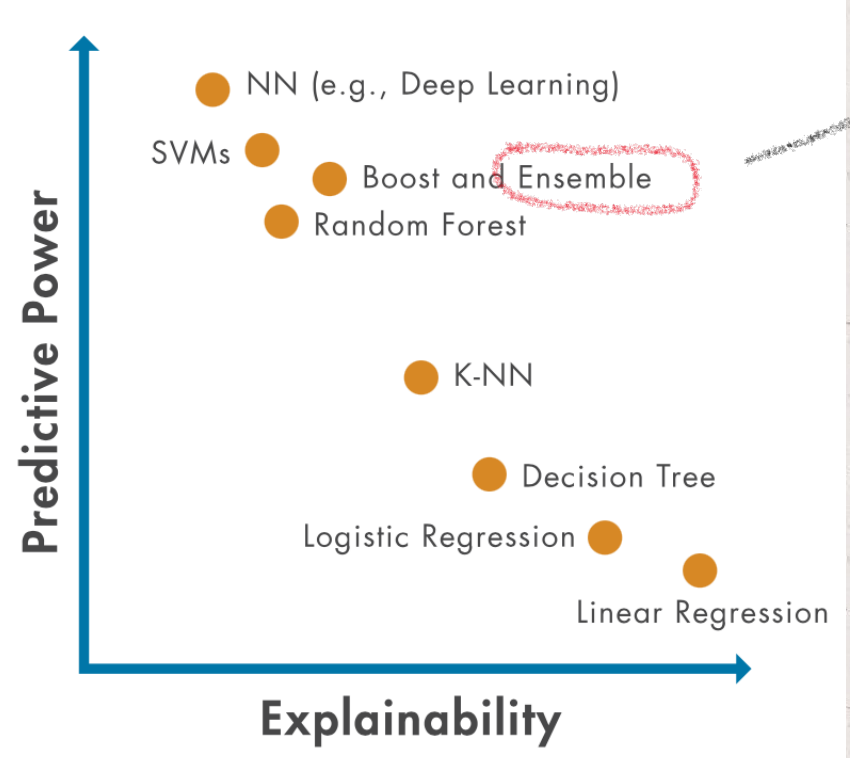
Motivation?



How to reduce the gap ?

Today's learning objectives

- KNN, Decision Trees, Ensembles

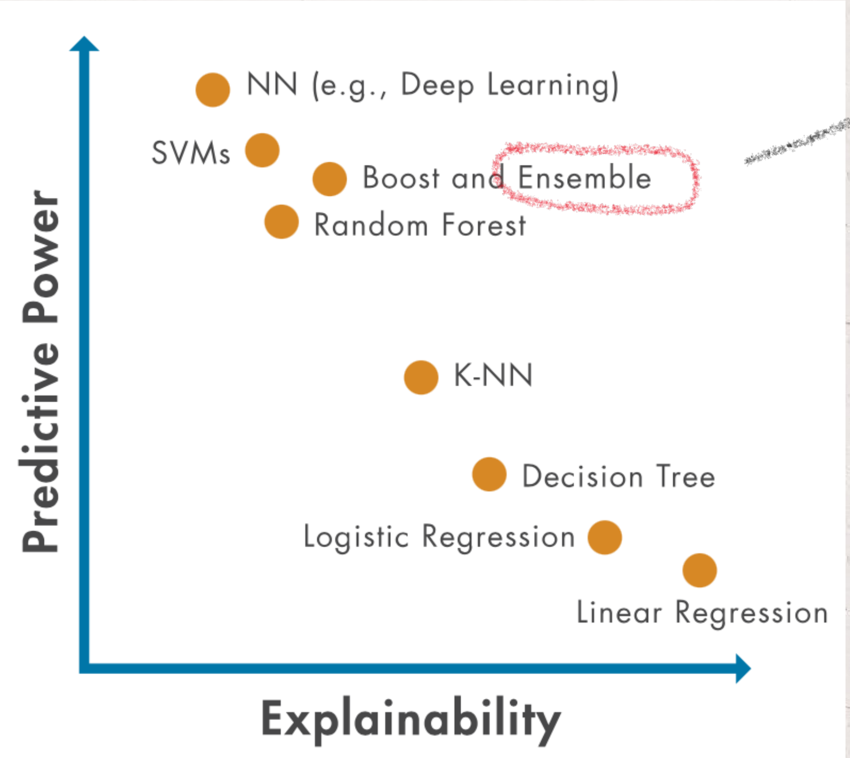


1. Train multiple independent models
2. At test time, average their results

Enjoy 2% extra performance (helps with regularization as well!)

Today's learning objectives

- KNN, Decision Trees, Ensembles



1. Train multiple independent models
2. At test time, average their results

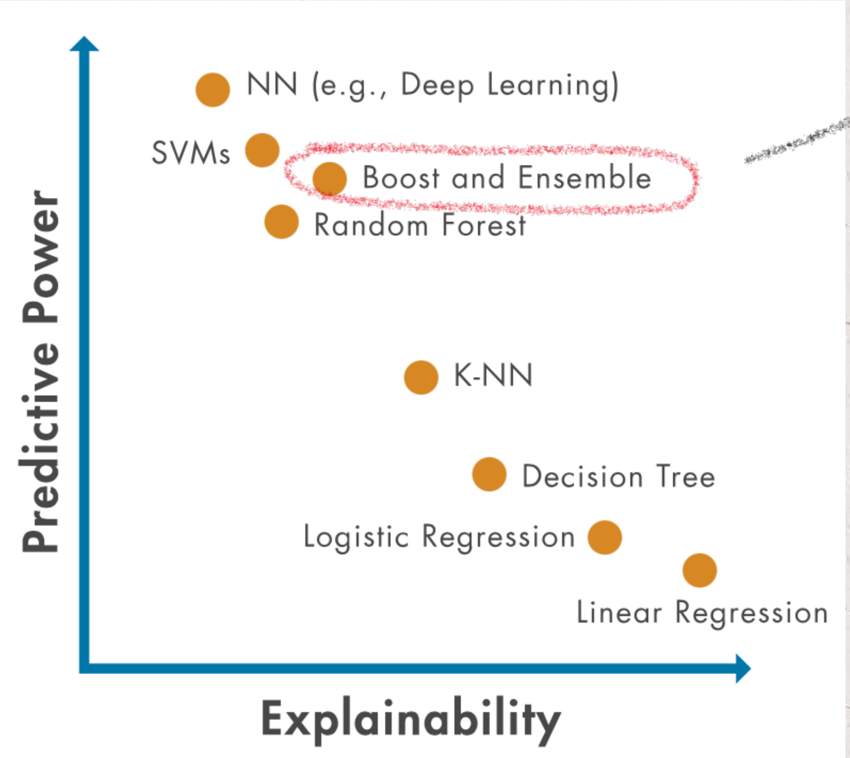
Enjoy 2% extra performance

Ensemble multiple deep networks?

computationally expensive?

Today's learning objectives

- KNN, Decision Trees, Ensembles



1. Train multiple independent models
2. At test time, average their results

Enjoy 2% extra performance

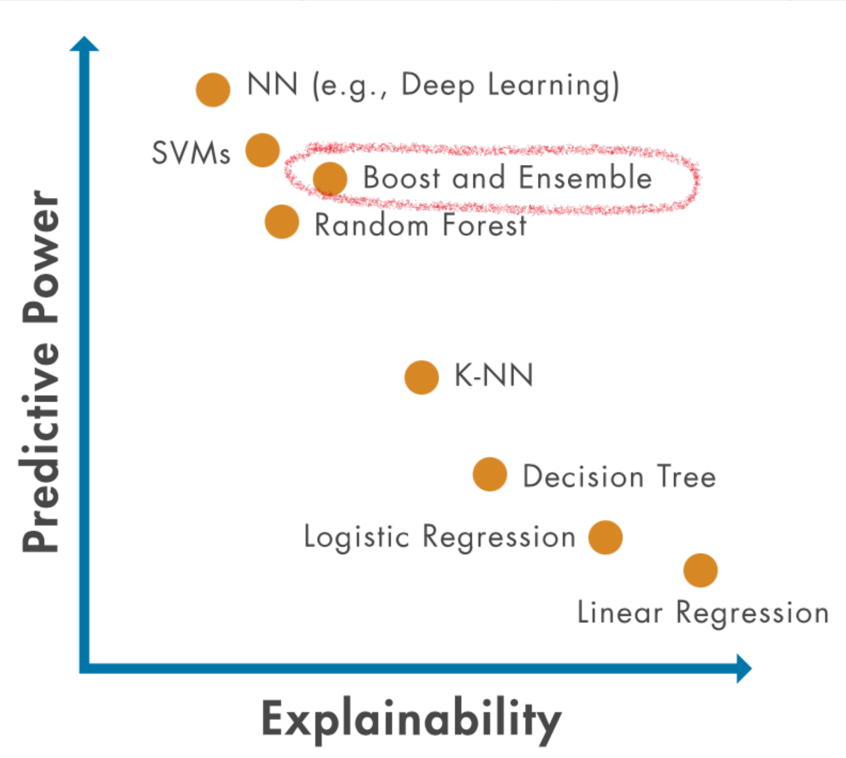
Ensemble multiple deep networks?

computationally expensive?

Solution 1: **Dropout** is equivalent to training a large ensemble of different models that share parameters.

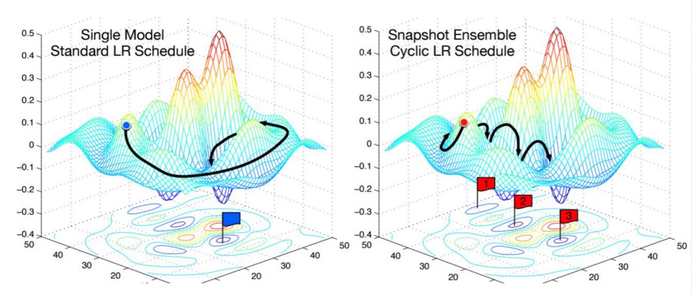
Today's learning objectives

- KNN, Decision Trees, Ensembles



leverage recent work on cyclic learning rate schedules

https://arxiv.org/abs/1704.00109



Loshchilov and Hutter, "SGDR: Stochastic gradient descent with restarts", arXiv 2016
Huang et al, "Snapshot ensembles: train 1, get M for free", ICLR 2017
Figures copyright Yixuan Li and Geoff Pleiss, 2017. Reproduced with permission.

a single neural network, converging to several local minima along its optimization path and saving the model parameters.

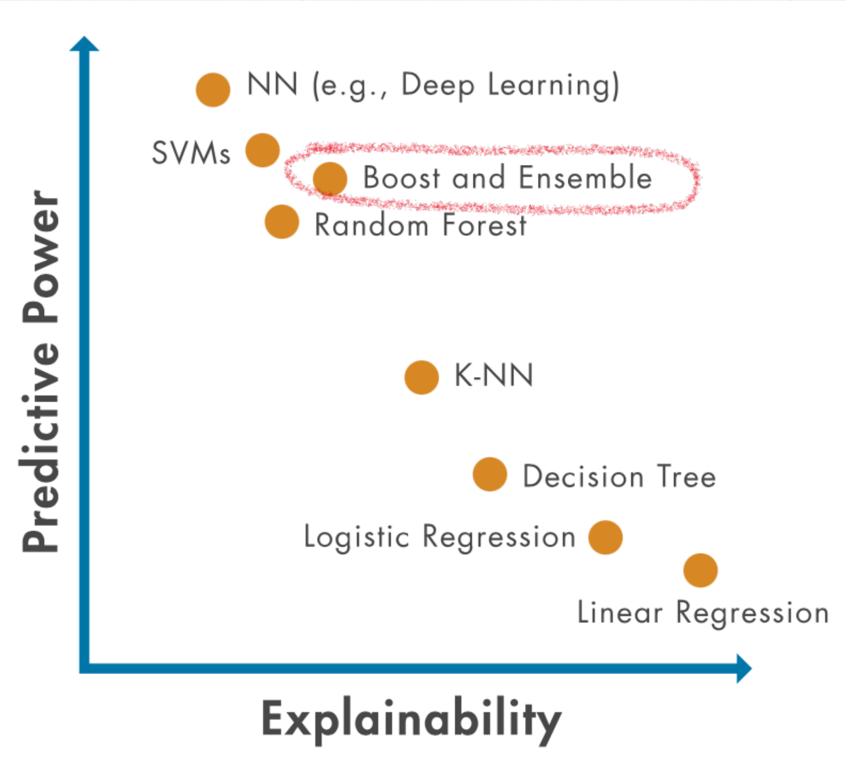
Ensemble multiple deep networks?

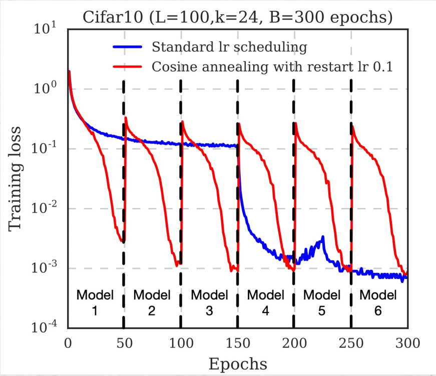
computationally expensive?

Solution 2: Snapshot Ensembles, instead of training independent models, use multiple snapshots of **a single model during training**

Today’s learning objectives

- KNN, Decision Trees, Ensembles

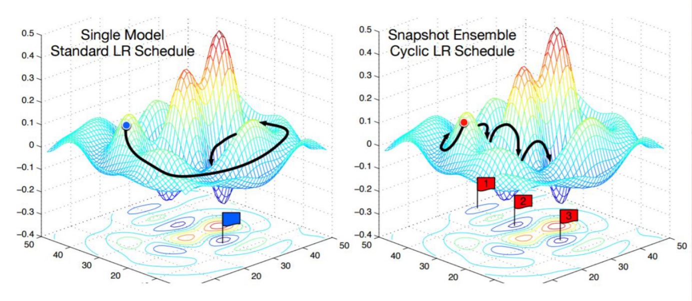




change learning rate aggressively during training (high-low-high, after high the new weights are likely to be better, so save them as a model snapshot. At the end average the results

leverage recent work on cyclic learning rate schedules

https://arxiv.org/abs/1704.00109



Loshchilov and Hutter, “SGDR: Stochastic gradient descent with restarts”, arXiv 2016
Huang et al, “Snapshot ensembles: train 1, get M for free”, ICLR 2017
Figures copyright Yixuan Li and Geoff Pleiss, 2017. Reproduced with permission.

a single neural network, converging to several local minima along its optimization path and saving the model parameters.

Ensemble multiple deep networks?

computationally expensive?

Solution 2: Snapshot Ensembles, instead of training independent models, use multiple snapshots of **a single model during training**