COMS30035, Machine learning: Combining Models 4, Conditional Mixture Models

Edwin Simpson
edwin.simpson@bristol.ac.uk

Department of Computer Science, SCEEM
University of Bristol

November 16, 2023

Agenda

- Model Selection
- Model Averaging
- Ensembles: Bagging
- Ensembles: Boosting and Stacking
- Tree-based Models
- Conditional Mixture Models
- Ensembles of Humans

Recap

- Model selection: choose the right model for the whole dataset –> *hard selection*
- Bayesian model averaging (BMA): probabilistically select the right model for the whole dataset –> *soft selection*
- Decision trees: split the feature space and model each region by one leaf node –> *hard selection depending on features*

Recap

- Model selection: choose the right model for the whole dataset –> *hard selection*
- Bayesian model averaging (BMA): probabilistically select the right model for the whole dataset –> *soft selection*
- Decision trees: split the feature space and model each region by one leaf node –> *hard selection depending on features*
- Conditional mixture models: perform a soft, probabilistic split of the feature space –> *soft selection depending on features*

Mixture of Experts

- Each data point is processed by a weighted combination of specialised 'expert' models
- Weights: probabilities that depend on the features $\boldsymbol{x}_n$ of the data point

Mixture of Experts

- Each data point is processed by a weighted combination of specialised 'expert' models
- Weights: probabilities that depend on the features $\boldsymbol{x}_n$ of the data point
- Think of medical diagnosis: based on the patient's symptoms, a GP refers the patient to a specialist

Mixture of Experts

- Each data point is processed by a weighted combination of specialised 'expert' models
- Weights: probabilities that depend on the features $\boldsymbol{x}_n$ of the data point
- Think of medical diagnosis: based on the patient's symptoms, a GP refers the patient to a specialist
- If they are unsure what is causing the symptoms, they may send the patient to multiple specialists for examination.

Mixture of Experts

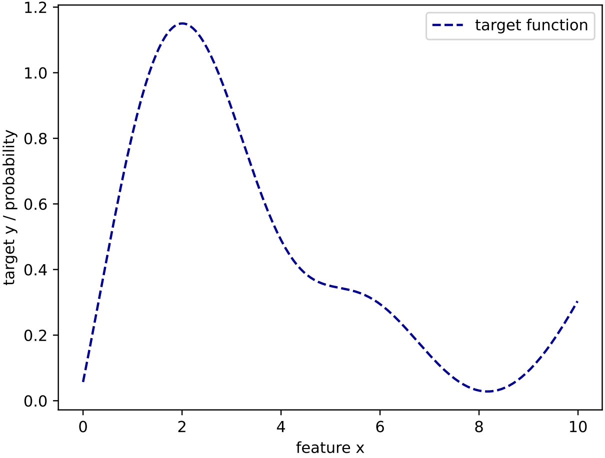
- Each data point is processed by a weighted combination of specialised 'expert' models
- Weights: probabilities that depend on the features $\boldsymbol{x}_n$ of the data point
- Think of medical diagnosis: based on the patient's symptoms, a GP refers the patient to a specialist
- If they are unsure what is causing the symptoms, they may send the patient to multiple specialists for examination.
- Similarly, some inputs $\boldsymbol{x}_n$ may require a combination of expert models

Mixture of Experts

- Each data point is processed by a weighted combination of specialised 'expert' models
- Weights: probabilities that depend on the features $\boldsymbol{x}_n$ of the data point
- Think of medical diagnosis: based on the patient's symptoms, a GP refers the patient to a specialist
- If they are unsure what is causing the symptoms, they may send the patient to multiple specialists for examination.
- Similarly, some inputs $\boldsymbol{x}_n$ may require a combination of expert models
- Contrast with decision trees, which assign each data point to a single leaf node

Mixture of Experts

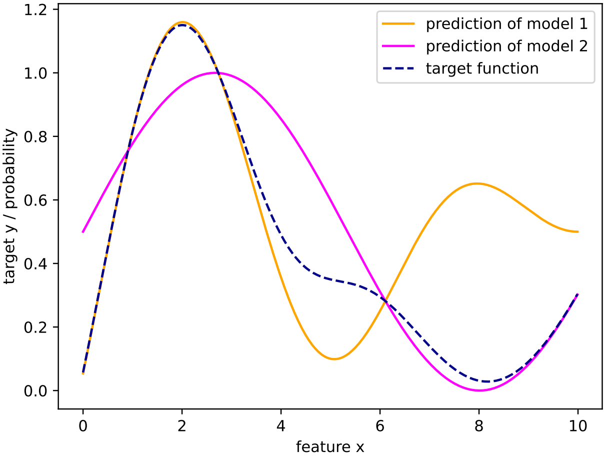
Imagine we would like to learn a model of this function:



Edwin Simpson
edwin.simpson@bristol.ac.uk

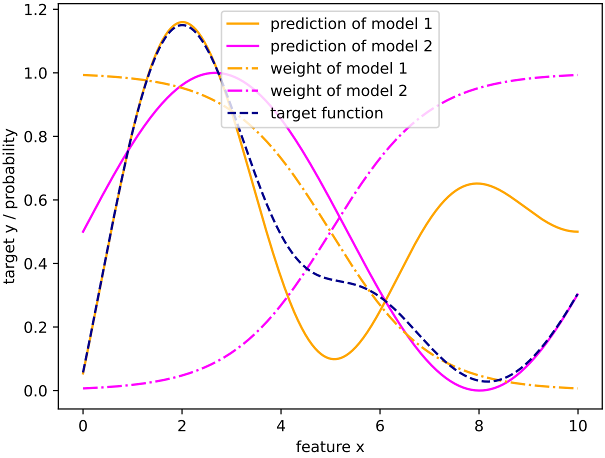
Mixture of Experts

We have two expert models: model 1 is close to our target on the left, model 2 is close to our target on the right.



Mixture of Experts (MoE)

MoE reproduces the target function by learning weights for each model and taking a weighted sum of their predictions.



Mixture of Experts

- Goal: predict target variable t_n given features $\boldsymbol{x}_n$
- Component distribution depends on input feature vector $\boldsymbol{x}_n$.

$$p(t_n|\boldsymbol{x}_n, \phi, \boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k(\boldsymbol{x}_n) p(t_n|\boldsymbol{x}_n, \phi_k) \quad (1)$$

- $\pi_k(\boldsymbol{x}_n)$ is the weight for model k in a combination of models.
- The weights can be learned as part of EM (see Bishop 14.5.3 for more)

Now do the quiz!

Please do the quiz for this lecture on Blackboard.