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 8, 2023

Agenda

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

Recap

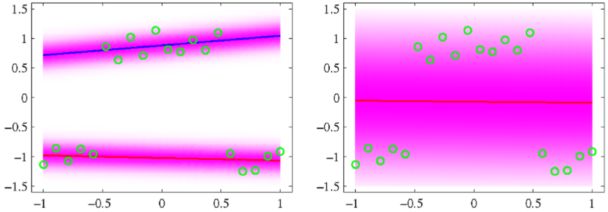
- Model selection: choose the right model for the dataset –> *hard selection*
- Bayesian model averaging (BMA): probabilistically select the right model for the dataset by weighting candidate models –> *soft selection*
- Decision trees: split the feature space so that each area is modelled by one leaf node –> *hard selection depending on features*
- Conditional mixture models: probabilistically *combine* models
- Mixture of experts: probabilistically weight the models depending on inputs –> *combine depending on features*

Conditional Mixture Models

- Similar to the mixture models we saw in earlier lectures
- For regression or classification, rather than clustering
  - Target variable t
  - Feature vector $\boldsymbol{x}$
  - Component density $\boldsymbol{\pi}$
  - Parameters of observation distribution ϕ
  - Goal: estimate $p(t|\boldsymbol{x}, \phi, \boldsymbol{\pi})$
  - Learn parameters using EM
- Each data point is generated from one mixture component as in the Gaussian mixture model (GMM).

Conditional Mixture Models

- Predictive posterior: $p(t_n|x_n, \phi, \boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k p(t_n|\boldsymbol{x}_n, \phi_k)$
- Type of distribution for $p(t_n|\boldsymbol{x}_n, \phi_k)$ depends on the data types of t_n and $\boldsymbol{x}_n$
- e.g., Gaussians for regression
- e.g., logistic model for classification.
- Allows us to model multimodal data:

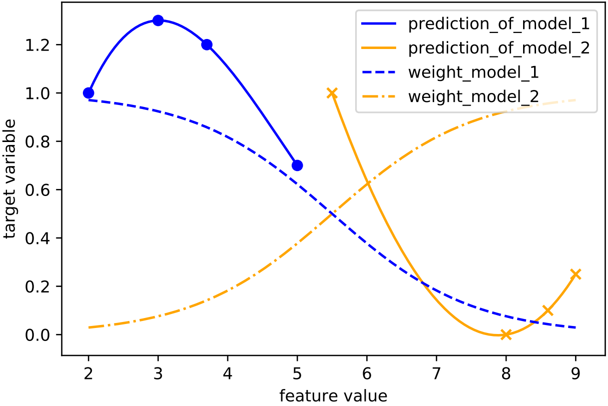


See Bishop (2006), 14.5.1 and 14.5.2.

Mixture of Experts

- Extends the conditional mixture model to weight the components differently for each data point depending on its features $\boldsymbol{x}_n$.
- Rather than a single, complex model, each part of the input space is dealt with by a specialised expert models.
- 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



Mixture of Experts

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

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

- $\pi_k(\boldsymbol{x}_n)$ acts like a weight for model k in a combination of models.
- The weights can also be learned using EM.

Now do the quiz!

Please do the quiz for this lecture on Blackboard.