CS-667 Advanced Machine Learning

Nazar Khan

PUCIT

Mixture Density Networks

Forward and Inverse Problems

- Goal of supervised learning: model conditional distribution $p(\mathbf{t}|\mathbf{x})$.
- For simple regression problems $p(\mathbf{t}|\mathbf{x})$ is assumed to be Gaussian.
- However, *practical machine learning problems* can have significantly non-Gaussian distributions.

Forward Problems

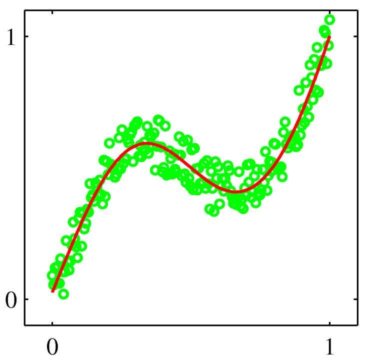

Figure: Successful neural network learning of a *uni-modal* forward problem ($t_n = x_n + 0.3\sin(2\pi x_n) + \epsilon$) using SSE function.

Inverse Problems

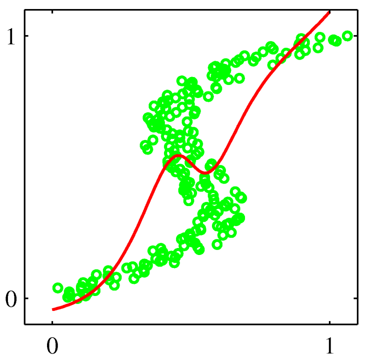

Figure: Unsuccessful neural network learning of a *multi-modal* inverse problem (roles of t_n and x_n reversed). *Reason for failure*: Training NN with SSE function implies $t \sim \mathcal{N}$. However, for multi-modal inverse problems $t \nsim \mathcal{N}$ and the learned model is a very poor fit of the underlying model.

Mixture Density Networks

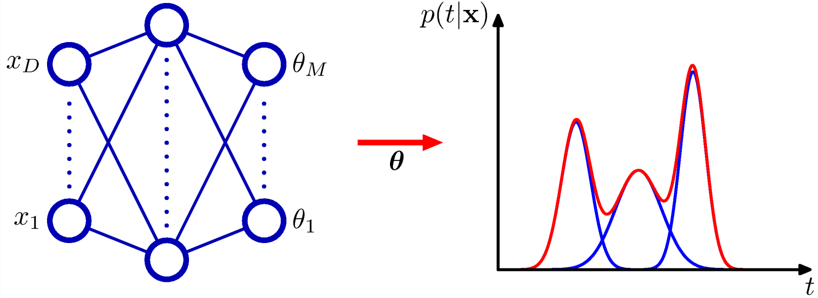

Figure: Mixture density network. Outputs are the mixture parameters $\boldsymbol{\theta}(\mathbf{x})$ corresponding to input $\mathbf{x}$. *Difference from earlier approaches*: Instead of learning parameters $\boldsymbol{\theta}$, we learn NN weights $\mathbf{w}$ that produce parameters $\boldsymbol{\theta}(\mathbf{x})$ that model the density conditioned on input $\mathbf{x}$.

Mixture Density Networks

$$p(\mathbf{t}|\mathbf{x}) = \sum_{k=1}^{K} \pi_k(\mathbf{x}) \mathcal{N}(\mathbf{t}|\boldsymbol{\mu}_k(\mathbf{x}), \sigma_k^2(\mathbf{x})\mathbf{I})$$

The component densities need not be Gaussian. They can be chosen according to the problem at hand (e.g Bernoulli densities if target t is a binary random variable).

Mixture Density Networks

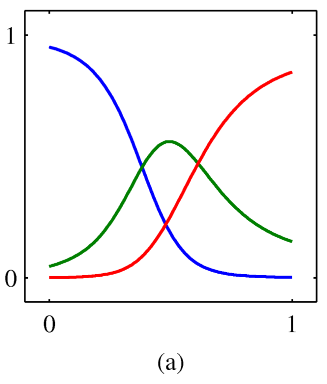

Figure: Mixing coefficients $\pi_k(x)$. At both small and large values of x where $p(t|x)$ is uni-modal, only one mixture component has a larger role. For intermediate values of x where the density is tri-modal, all 3 mixing coefficients have comparable values.

Mixture Density Networks

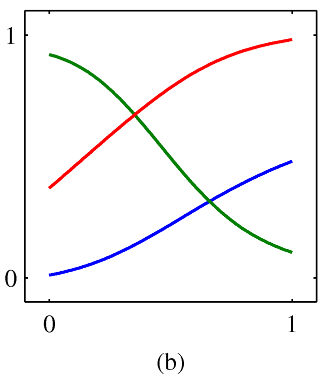

Figure: Means $\mu_k(x)$.

Mixture Density Networks

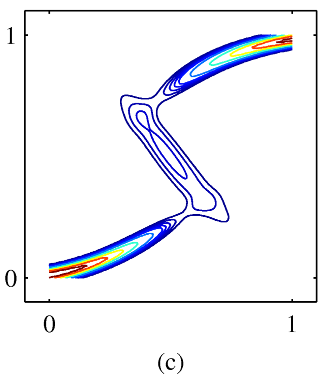

Figure: Contours of $p(t|x)$

.

Mixture Density Networks

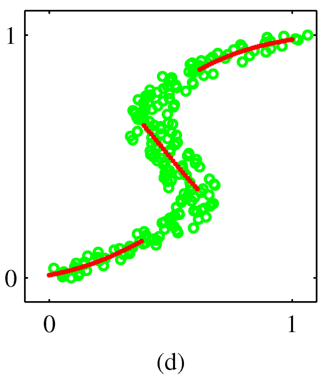

Figure: Approximate modes of conditional density $p(t|x)$.

Assignment 7

EM for Gaussian Mixture Model

- Density estimation via Gaussian Mixture Model (GMM).
  - Code up a generic implementation of learning a GMM via the EM algorithm in function `[mixing_coefs,means,covariance_mats]=learn_gmm(X,K)` where `X` is a $D \times N$ data matrix and `K` is the number of Gaussian components.
  - Regenerate Figure 9.8 in Bishop's book.
- Submit your_roll_number_GMM.zip containing
  - code,
  - generated image, and
  - report.txt/pdf explaining your results.
- Due Thursday, May 18, 2017 before 5:30 pm on \\printsrv.