Intro to Density Estimators

Density estimation – attempt to recover the probability distribution function from the data. Simplest example is histogram.

Kernel Density Estimation (KDE) – most common method of estimating density. Given *n* i.i.d. data points, the estimated density is:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

Where *K* is the Kernel function (itself a unimodal pdf with mean 0, often chosen to be Gaussian – there is one called the Epanechnikov which is minimal MSE) and *h* is the bandwidth – intuitively we want this to be as small as possible, but there is a bias-variance trade off, and too small a bandwidth is undersmoothed and the density function is overfitted. Silverman's rule is often used as rule of thumb.

Neural Density Estimation – using neural networks to do density estimation. Very useful for high dimensional data, intricate data structures or for use in Bayesian inference. Some techniques are:

- Autoregressive models – decompose multivariate density into chain rule product of conditional densities i.e. $p(x)=p(x_1)p(x_2|x_1)p(x_3|x_1,x_2)\ldots p(x_d|x_1,x_2,\ldots,x_{d-1})$. Each conditional probability is modelled using neural networks, for example masking methods
- *Normalizing flows* - taking a simple probability distribution and applying a series of invertible, differentiable transformations to make the pdf more complex. MAF is masked autoregressive flow, uses autoregression to define the transformation in the flow
- Variational Autoencoders – VAEs learn the distribution of lower dimensional latent variables, and assume they generate the data, and you can model the conditional distribution of the data with these variables; the encoder is what approximates the posterior which is used in the conditional distribution