Fast Kernel Smoothing in Projection Pursuit

David Hofmeyr
Dept. Statistics and Actuarial Science

October 25, 2019

Kernel Smoothing and Projection Pursuit

- Kernel Smoothing: Non-parametric function estimation through locally weighted averages,

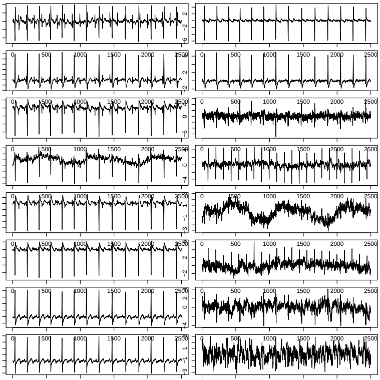
$$\hat{f}(x) = \sum_{i=1}^{n} K\left(\frac{|x - x_i|}{h}\right)\omega_i$$

  - $\mathcal{O}(nm)$ to evaluate (directly) at m points
  - If $K(x) = \text{poly}(x)e^{-x}$, then exact evaluation in $\mathcal{O}(n\log(n) + m)$[1]
- Proj. Pursuit: Find $V \in \mathbb{R}^{p \times p'}$ to maximise some functional of the density/distribution of $X'V$ (or conditional $Y|X'V$)
  - We don't know the distribution of X, so estimate that of $X'V$ with kernels

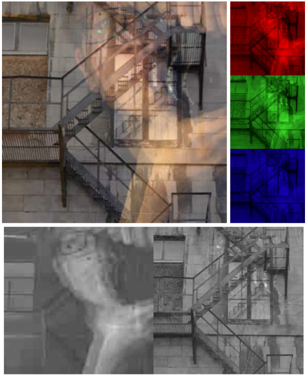
[1] Check my github (soon) for code

Independent Component Analysis[2]

Identify independent (latent) sources by minimising KL divergence between $f_{X'V}$ and $\prod_{i=1}^{p'} f_{X'V_i}$ $\iff$ minimise the sum of entropies of $X'V_i$'s



(a) Foetal ECG



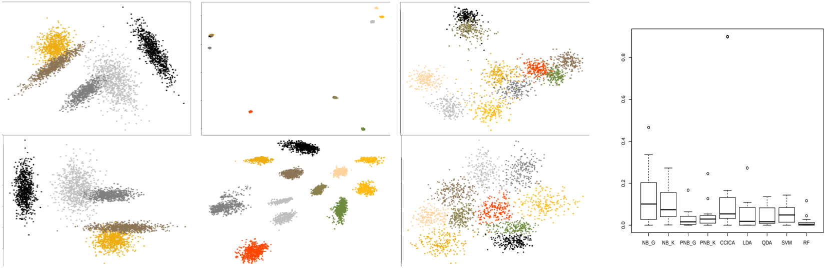
(b) Reflection removal

[2] joint with HP Bakker, F Kamper, M Melonas. ECG data from de Lathauwer et al., " Fetal electrocardiogram extraction by blind source subspace separation", IEEE TBE, 2000. Image from Shih et al., "Reflection removal using ghosting cues", CVPR, 2015

Optimal Projections for Naïve Bayes[3]

NB: class conditional independence $\Rightarrow$ (potentially) heavy bias $\Rightarrow$ find a projection under which this assumption is more plausible

$$\max_V \prod_{i=1}^{n} P(Y_i = y_i | X'V) = \max_V \prod_{i=1}^{n} \frac{\hat{f}_{X'V}(x_i'V|y_i)\pi_{y_i}}{\sum_k \hat{f}_{X'V}(x_i'V|k)\pi_k}$$



(c) Simul. (d) Yale faces B (e) Digits (f) Performance

[3] joint with M Melonas. Yale B: `http://vision.ucsd.edu/~leekc/ExtYaleDatabase/`, Digits: `https://archive.ics.uci.edu/ml`

Projection Pursuit Regression for DSM[4]

PPR: "like a single layer NN with non-parametric activation function"

$$\hat{f}(x) = \mu + \sum_{i=1}^{k} \alpha_i \hat{f}_i(x' V_i)$$

Can we ignore explicit spatial variation by using flexible regressors?

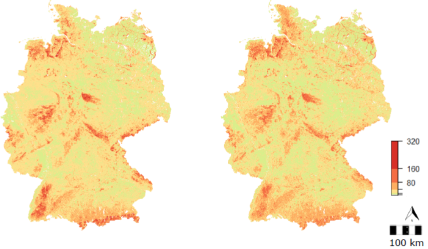


Figure 2. SOC maps of Germany using the predictions of the RK model (left) and the PPR model (right).

[4] joint with S Van der Westhuizen, G Heuvelnik, L Poggio. Data from ISRIC – World Soil Information.

Other Interests

- Model selection and estimating generalisation performance
- Clustering
- Semi-supervised learning
- Asymptotics for non-parametrics (mostly kernel type)

Please feel free to come and chat if you're interested in any of these topics