Stochastic Linear Regression

$$\eta(x) = \sum_{k=1}^{K} \beta^{(k)} h_k(x), \quad \beta^{(k)} \sim \mu_k \text{ indep}$$

(or overall, $(\beta^{(1)} \dots \beta^{(K)}) \sim \mu$)

Example $\eta(x) = \beta^{(1)} + \beta^{(2)} x, \quad \beta^{(k)} \sim \mu_k$

$$h_1(x) \equiv 1, \quad h_2(x) = x$$

$$f_{\eta(x)}(y) = \int f_1(y - b_2 x) f_2(b_2)\, db_2, \text{ see } (*) \text{ below}$$

Denote by $F_k(b) = P\{\beta^{(k)} \le b\} = \mu_k(-\infty, b]$ – c.d.f.'s

Assume that h_k are continuous and μ_k has a density $f_k(b) = \frac{d}{db} F_k(b)$ for all k.

Then $\eta(x)$ has also a density $f_{\eta(x)}(y)$:

$$F_{\eta(x)}(y) = P\{\eta(x) \le y\} = P\Big\{\sum_{k=1}^{K} \beta^{(k)} h_k(x) \le y\Big\}$$

$$= E\, P\Big\{\beta^{(1)} \le \frac{1}{h_1(x)}\Big(y - \sum_{k=2}^{K} \beta^{(k)} h_k(x)\Big) \Big| \beta_2, \dots, \beta_K\Big\}$$

$$= \int \cdots \int_{\mathbb{R}^{K-1}} F_1\Big(\frac{1}{h_1(x)}\Big(y - \sum_{k=2}^{K} b_k h_k(x)\Big)\Big) f_2(b_2) \dots f_K(b_K)\, db_2 \dots db_K$$

hence $f_{\eta(x)}(y) = \frac{d}{dy} F_{\eta(x)}(y) =$

$$= \frac{1}{h_1(x)} \int_{\mathbb{R}^{k-1}} \!\!\dots\! \int f_1\left(\frac{1}{h_1(x)}\left(y - \sum_{k=2}^{K} b_k h_k(x)\right)\right) f_2(b_2)\dots f_K(b_K)\, db_2 \dots db_K$$

$$= \frac{1}{h_1(x)} \mathbb{E}\, f_1\left(\frac{1}{h_1(x)}\left(y - \sum_{k=2}^{K} \beta^{(k)} h_k(x)\right)\right) \quad (*)$$

which can be estimated by Monte-Carlo

Given observations (x_i, y_i), $i=1,\dots,N$

$$\mathrm{LogLik}(\bar{y}) = \sum_{i=1}^{N} \log f_{\eta(x_i)}(y_i) \longrightarrow \text{max over } f_1 \dots f_K$$

Gaussian linear regression

$$\beta^{(k)} \sim N(m_k, V_k)$$

$$V_k = \operatorname{var} \beta^{(k)}$$

Then $\eta(x) \sim N(m(x), V(x))$, where

$$m(x) = E\,\eta(x) = \sum_{k=1}^{K} m_k h_k(x)$$

$$V(x) = \operatorname{var} \eta(x) = \sum_{k=1}^{K} V_k(x) h_k^2(x)$$

by independence of $\beta^{(k)}$'s

Given observations $(x_i, y_i), i=1,\dots,N$

$$\mathrm{LogLik}(\bar{y}) = \sum_{i=1}^{N} \log f_{\eta(x_i)}(y_i) =$$

$$= \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi V(x_i)}} e^{-\frac{(y_i - m(x_i))^2}{2V(x_i)}}$$

$$= c - \sum_i \left[\frac{1}{2}\log V(x_i) + \frac{(y_i - m(x_i))^2}{2V(x_i)}\right]$$

We have $\dfrac{\partial m(x)}{\partial V_k} = \dfrac{\partial V(x)}{\partial m_k} = 0$

$$\frac{\partial m(x)}{\partial m_k} = h_k(x) \qquad \frac{\partial V(x)}{\partial V_k} = h_k^2(x)$$

$$\frac{\partial \mathrm{LogLik}(\bar{y})}{\partial m_k} = \sum_i \frac{h_k(x_i)(y_i - m(x_i))}{V(x_i)} = 0$$

$$\frac{\partial \mathrm{LogLik}(\bar{y})}{\partial V_k} = -\frac{1}{2}\sum_i \frac{h_k^2(x_i)}{V(x_i)} + \frac{1}{2}\sum_i \frac{h_k^2(x_i)(y_i - m(x_i))^2}{V^2(x_i)}$$

$$= \frac{1}{2}\sum_i \frac{h_k^2(x_i)}{V^2(x_i)}\left[(y_i - m(x_i))^2 - V(x_i)\right] = 0$$

Solved numerically!