5.3

(a) $$E_X[\hat{p}(x)] = \sum_i E_{x_i}\left[\frac{1}{n}\sum_i k(x-x_i)\right]$$
$$= \sum_i \int \frac{1}{n}\sum_i k(x-x_i)\cdot p(x_i)\cdot dx_i$$
$$= \frac{1}{n}\sum_i \int \mathcal{N}(x-x_i|0,h^2)\cdot\mathcal{N}(x_i|\mu,\alpha^2)\,dx_i$$
$$= \frac{1}{n}\sum_i \int \mathcal{N}(x_i|x,h^2)\mathcal{N}(x_i|\mu,\alpha^2)\,dx_i$$
$$= \frac{1}{n}\sum_i \int \mathcal{N}(x|\mu,\alpha^2+h^2)\,\mathcal{N}\left(x_i|c\left(\frac{x}{h^2}+\frac{\mu}{\alpha^2}\right),\frac{1}{\frac{1}{\alpha^2}+\frac{1}{h^2}}\right)\cdot dx_i$$
$$= \mathcal{N}(x|\mu,\alpha^2+h^2)$$

(b) $$\mathrm{Var}_X[\hat{p}(x)] = \mathrm{Var}\left(\sum_i \frac{1}{n}\tilde{k}(x-x_i)\right)$$
$$= \frac{1}{n}\mathrm{Var}_z[\tilde{k}(x-z)]$$
$$= \frac{1}{n}\left(E_z(\tilde{k}(x-z)^2) - \mathcal{N}^2(x|\mu,\alpha^2+h^2)\right]$$
$$= \frac{1}{n}\left(\int \mathcal{N}(x-z|0,h^2)^2 p(z)\,dz - \mathcal{N}^2(x|\mu,\alpha^2+h^2)\right]$$
$$= \frac{1}{n}\left(\underbrace{\int \frac{1}{2\pi h^2}\cdot e^{-\frac{1}{2}\frac{2(x-z)^2}{h^2}}p(z)\,dz}\; - \mathcal{N}(x|\mu,\alpha^2+h^2)^2\right]$$
$$\left(\frac{1}{2\sqrt{\pi}h}\frac{1}{\sqrt{2\pi}\cdot\frac{h}{\sqrt{2}}}\,e^{-\frac{1}{2}\frac{(x-z)^2}{(\frac{h}{\sqrt{2}})^2}}p(z)\right) = \frac{1}{2\sqrt{\pi}h}\mathcal{N}\left(z|x,\frac{h^2}{2}\right)\cdot\mathcal{N}(z|\mu,\alpha^2)$$
$$= \frac{1}{2\sqrt{\pi}h}\mathcal{N}(\;)\,\mathcal{N}\left(x|\mu,\alpha^2+\frac{h^2}{2}\right)$$
$$\approx \frac{1}{2\sqrt{\pi}h}\mathcal{N}(\;)\,\mathcal{N}(x|\mu,\alpha^2)$$
$$\frac{1}{2\sqrt{\pi}h}\cdot p(x)$$
$$= \frac{1}{2nh\sqrt{\pi}}\left(p(x) - 2nh\sqrt{\pi}\,\mathcal{N}(x|\mu,\alpha^2+h^2)^2\right]$$
$$\approx \frac{1}{2nh\sqrt{\pi}}p(x)$$

(c) $$\frac{1}{\sqrt{2\pi}\alpha}e^{-\frac{(x-\mu)^2}{2\alpha^2}} - \frac{1}{\sqrt{2\pi(\alpha^2+h^2)}}\cdot e^{-\frac{(x-\mu)^2}{2(\alpha^2+h^2)}}$$
$$= f(\alpha^2) - f(\alpha^2+h^2)$$
$$f(u) = \frac{1}{\sqrt{2\pi u}}e^{-\frac{(x-u)^2}{2u}}$$
$$\frac{df(u)}{du} = -\frac{1}{2\sqrt{2\pi}u^{\frac{3}{2}}}e^{-\frac{(x-\mu)^2}{2u}} + \frac{1}{\sqrt{2\pi u}}e^{-\frac{(x-\mu)^2}{2u}}\cdot\frac{(x-\mu)^2}{2u^2}$$
$$= p(x)\left(-\frac{1}{2u}+\frac{(x-\mu)^2}{2u^2}\right)$$

$$x \qquad - h^2 \frac{d f(\alpha^2)}{d\alpha^2}$$

$$= p(x) \left(\qquad \frac{h^2}{2\alpha^2} - \frac{(x-\mu)^2}{2\alpha^4} \right)$$

$$= \frac{h^2}{2\alpha^2} \left(1 - \frac{(x-\mu)^2}{\alpha^2}\right) \cdot p(x))$$

(d) V

5-4

(a)

$$\hat{p}(x) = \frac{1}{nh} \cdot \sum_i k\left(\frac{x-x_i}{h}\right)$$

$$\hat{E}[\hat{p}(x)] = \frac{1}{n} \sum_i E_z \tilde{k}(x-z)$$

$$= \frac{1}{na} \sum_i \int_0^a \frac{1}{h} \cdot e^{-\frac{x-z}{h}} \cdot \quad dz$$

① $$= \frac{1}{an} \sum_i \int_0^a e^{\frac{z-x}{h}} \, d\,\frac{z-x}{h} \quad (\text{if } z \le a \le x)$$

$$= \frac{1}{a} \left(e^{\frac{z-x}{h}}\Big|_a - e^{\frac{z-x}{h}}\Big|_0 \right)$$

$$= \frac{1}{a}(e^{a/h} - 1) e^{-x/h}.$$

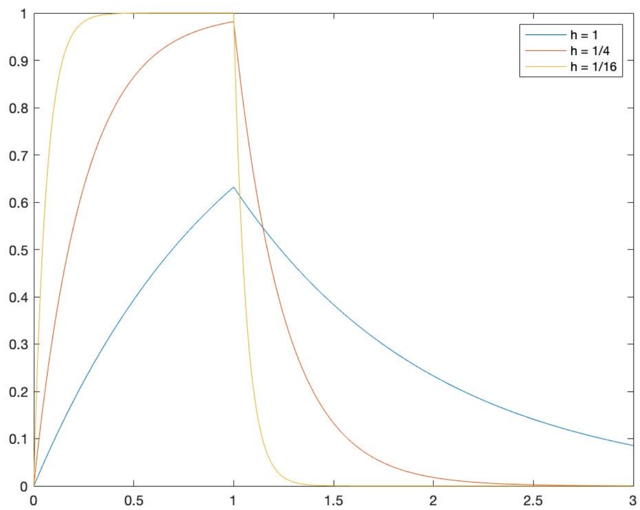
② $$= \frac{1}{an} \sum_i \int_0^x e^{\frac{z-x}{h}} \, d\,\frac{z-x}{h}$$

$$= \frac{1}{a}\left(e^{\frac{z-x}{h}}\Big|_x - e^{\frac{z-x}{h}}\Big|_0 \right) \quad (\text{if } a \ge x)$$

$$= \frac{1}{a}\left(1 - e^{-\frac{x}{h}}\right)$$

③ $$0 \qquad\qquad x < 0$$

(b)



(c)

$$\text{Bias} = \left|\frac{1}{a}\left(1-e^{-\frac{x}{h}}\right) - \frac{1}{a}\right|$$

$$= \left|-\frac{e^{-\frac{x}{h}}}{a}\right| = \frac{e^{-\frac{x}{h}}}{a}$$

we need $\frac{e^{-\frac{x}{h}}}{a} \le 0.01$

for all $x \ge 0.01a$.

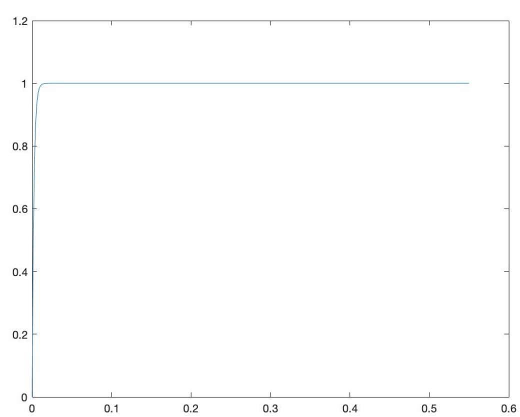
$$e^{-\frac{x}{h}} \le 0.01a.$$

$$-\frac{x}{h} \le -2\ln 10 + \ln a.$$

$$\frac{x}{h} \ge 2\ln 10 - \ln a$$

$$h \ge \frac{0.01a}{2\ln 10 - \ln a}.$$

(d) $h \ge \frac{0.01}{2\ln 10}$ if $a = 1$



5.5 (a) $\bar{k}(r) = \max\{0, 1-r\}$, yes

(b) $\bar{g}(r) = \begin{cases} 1, & r \le 1 \\ 0, & r > 1 \end{cases}$

(c) $$\hat{x}^{(k+1)} = \frac{\sum_i x_i \bar{g}(\|\frac{x-x_i}{h}\|^2)}{\sum_i \bar{g}(\|\frac{x-x_i}{h}\|)}$$

$$= \frac{\sum_i x_i \cdot \text{ind}(x-x_i)}{\sum_i \text{ind}(x-x_i)}, \quad \text{where } \text{ind}(x-x_i) = \begin{cases} 1, & -h \le x-x_i \le h \\ 0, & \text{else} \end{cases}$$

(d) mean-shift using gaussian

$$\hat{x}^{(k+1)} = \frac{\sum_i x_i e^{-\frac{1}{2}\|\frac{x-x_i}{h}\|^2}}{\sum_i e^{-\frac{1}{2}\|\frac{x-x_i}{h}\|^2}}$$

(exponentially decay.)

consider all x_i, though give low weight to far ones

K-means: iterate ⇒ assign to clusters and compute means.

5.6

(a) $\nabla p(x) =$

$$\sum_j \pi_j \nabla \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_j|^{d}} e^{-\frac{1}{2}\|x-\mu_j\|_{\Sigma_j}}$$

$$= \sum_j \pi_j \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|_{\Sigma_j}} \cdot -\frac{1}{2}$$

$$\nabla\left((x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j) \right)$$

$$\left(= \nabla\left(x^T \Sigma_j^{-1} x - 2x^T \Sigma_j^{-1} \mu_j + \mu_j^T \Sigma_j^{-1} \mu_j \right) \right)$$

$$= 2\Sigma_j^{-1} x - 2\Sigma_j^{-1} \mu_j$$

consider ∂x_i

$$x^T A x = \left[\sum_i x_i a_{i1}, \dots, \sum_i x_i a_{in}\right] x$$

$$\left(\begin{aligned} &= \sum_i \sum_j x_i a_{ij} x_j \\ \frac{\partial}{\partial x_i} &= \sum_i a_{il} x_i + \sum_j a_{ij} x_j \\ &\Rightarrow Ax + A^T x \end{aligned}\right)$$

$$= \sum_j \pi_j \frac{1}{(2d)^{\frac{d}{2}} |\Sigma_j|^{\frac{d}{2}}} \cdot e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j} \cdot -\frac{1}{2}\left(2\Sigma_j^{-1} x - 2\Sigma_j^{-1}\mu_j\right)$$

$$= \frac{\alpha}{(2d)^{\frac{d}{2}}} \cdot \sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j} \cdot \Sigma_j^{-1} \left(\frac{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j} \cdot \Sigma_j^{-1} \mu_j}{\underbrace{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j} \cdot \Sigma_j^{-1}}_{\text{(inverse)}}} - x \right)$$

$$= \underbrace{\frac{\alpha}{|2d|^{\frac{d}{2}}} \sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j}}_{\Downarrow \ \frac{1}{\lambda}} \left(\frac{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} \cdot e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j} \cdot \Sigma_j^{-1} \mu_j}{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j}} - \frac{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j} \cdot \Sigma_j^{-1}}{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j}} \cdot x \right)$$

gradient ascent

$$x^{(t+1)} = x^{(t)} + \lambda \nabla p(x^{(t)})$$

$$= \frac{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} \cdot e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j} \cdot \Sigma_j^{-1} \mu_j}{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j}} + \left(I - \frac{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j} \cdot \Sigma_j^{-1}}{\sum_j \pi_j \frac{1}{|\Sigma_j|^{\frac{d}{2}}} e^{-\frac{1}{2}\|x-\mu_j\|\Sigma_j}} \right) x^{(t)}$$

PS 5.2

(a) $\int \frac{1}{n}\sum_i \tilde{k}(x-x_i)\, x\, dx$

$= \frac{1}{n}\int \sum_i \tilde{k}(x-x_i)\cdot x \cdot dx$

$= \sum_i \frac{1}{n}\int \tilde{k}(x-x_i)(x-x_i)\, d(x-x_i)$
$\quad + \sum_i \frac{1}{n}\int \tilde{k}(x-x_i)\cdot x_i \cdot d(x-x_i)$

$= \dfrac{\sum_i x_i}{n}$

(b) $\hat{\Sigma} = \int \frac{1}{n}\sum_i \tilde{k}(x-x_i)(x-\hat{u})(x-\hat{u})^T dx$

$= \int \frac{1}{n}\sum_i \tilde{k}(x-x_i)\,(xx^T - 2x\hat{u}^T + \hat{u}\hat{u}^T)\cdot dx$

$= \int \frac{1}{n}\sum_i \tilde{k}(x-x_i)\,(xx^T - 2xx_i^T + x_ix_i^T + \hat{u}\hat{u}^T - x_ix_i^T)dx$

$= \int \frac{1}{n}\sum_i \tilde{k}(x-x_i)(x-x_i)(x-x_i)^T dx$

$+ \int \frac{1}{n}\sum_i \tilde{k}(x-x_i)\,(\hat{u}\hat{u}^T + x_ix_i^T - 2x_ix_i^T)\cdot dx$

$= H + \frac{1}{n}\sum_i (x_ix_i^T - 2\hat{u}x_i^T + \hat{u}\hat{u}^T)$

$= H + \frac{1}{n}\sum_i (x_i-\hat{u})(x_i-\hat{u})^T$

(c) From 5.1, the kernal density estimate $\hat{p}(x)$ is blurred by convolution.
From 5.2, we see that this blurring effect will not change the mean (as $\hat{u} = \frac{\sum_i x_i}{n}$), but increase the covarince by H. (as $\hat{\Sigma} = H + \frac{1}{n}\sum_i (x_i-\hat{u})(x_i-\hat{u})^T$).