SDE exercises p.141

① $H(\underline{x}) \hat{=} f(\underline{y}) = y_1 + 2y_2 + 2y_3 + 3 = \underline{w}^T \underline{y} + 3$

margin $m = \frac{1}{\|\underline{w}\|}$; $\underline{w} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

$\Rightarrow m = \frac{1}{\sqrt{\langle \underline{w}, \underline{w} \rangle}} = \frac{1}{\sqrt{1^2 + 2^2 + 2^2}} = \underline{\underline{\frac{1}{3}}}$

For support vectors :

$t_e(\underline{w}^T \underline{y}_e + b) = 1$

$A: (-1)\left[(-1) + 2\cdot(-1) + 2\cdot(-2) + 3\right] = 4$

$B: (-1)\left[2 + 2\cdot(-2) + 2\cdot(-1) + 3\right] = 1$ } support vectors

$C: (1)\left[-2 + 2\cdot(-2) + 2\cdot(2) + 3\right] = 1$

$\begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}$ with label $t_e = -1$

and $\begin{bmatrix} -2 \\ -2 \\ 2 \end{bmatrix}$ with label $t_e = +1$

---

② (a) $\boxed{\text{Dot products are semi-positive definite kernels}}$

To be shown: Let $k(x, y)$ be a dot product. Then, $k(x, y)$ is a positive definite kernel.

Let $k(x, y)$ be a dot product. That implies that $\exists\ \Phi$ and some dot product $\langle \cdot , \cdot \rangle$ such that $k(x,y) = \langle \Phi(x), \Phi(y) \rangle$.

Now pick any $\ell \in \mathbb{N}$ and any sequence $\{x_1, x_2, \dots, x_\ell\}$ and let $\underline{\underline{K}}$ the associated Gram matrix

$$\underline{\underline{K}} = \begin{bmatrix} & \vdots & \\ \cdots & k(x_i, x_j) & \cdots \\ & \vdots & \end{bmatrix}.$$

Then, $\forall\ \underline{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_\ell \end{bmatrix} \neq 0$

$$\Rightarrow \underline{c}^T \underline{\underline{K}}\, \underline{c} = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} c_i c_j k(x_i, x_j)$$

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} c_i c_j \langle \Phi(x_i), \Phi(x_j) \rangle$$ ... k is dot product

$$= \left\langle \sum_{i=1}^{\ell} c_i \Phi(x_i), \sum_{j=1}^{\ell} c_j \Phi(x_j) \right\rangle$$ ... inner product is a bilinear form

$$= \left\| \sum_{i=1}^{\ell} c_i \Phi(x_i) \right\|^2$$ ... inner product induces a norm

$$\geq 0$$ ... semi-positive definiteness

(b) Semi-positive definite kernels are dot product kernels

To be shown: Let $k(x,y)$ for $x,y \in \Omega$ be a semi-positive definite kernel. Then, $k(x,y)$ is a dot-product kernel.

$\Rightarrow$ cf. p. 61 theorem 3.11 in EECS Berkeley PDF (moodle)

(3) Show that $k(\underline{y}, \underline{y}') = (\underline{y}^T \underline{y}')^D$ with $D \in \mathbb{N}$ corresponds to an inner product in the feature space spanned by all possible Dth degree monomials

$\Rightarrow$ cf. EECS Berkeley PDF on embeddings corresponding to kernel constructions on p. 77 (p. 31 in the PDF).

Here, consider the core of it:

$$\langle \Phi(\underline{y}), \Phi(\underline{y}') \rangle = k(\underline{y}, \underline{y}')$$

$$= \sum_{m_1=1}^{K} \sum_{m_2=1}^{K} \cdots \sum_{m_D=1}^{K} (y_{m_1} \cdot y_{m_2} \cdot \ldots \cdot y_{m_D})(y'_{m_1} \cdot y'_{m_2} \cdot \ldots \cdot y'_{m_D})$$

$$= \left(\sum_{m_1=1}^{K} y_{m_1} y'_{m_1}\right) \cdot \left(\sum_{m_2=1}^{K} y_{m_2} y'_{m_2}\right) \cdot \ldots \cdot \left(\sum_{m_D=1}^{K} y_{m_D} y'_{m_D}\right)$$

$$= \left(\sum_{m=1}^{K} y_m y'_m\right)^D = \left(\langle \underline{y}, \underline{y}' \rangle\right)^D = (\underline{y}^T \underline{y}')^D$$

Example: $\quad D = 2, \quad K = 5$

$$\sum_{m_1=1}^{5} \sum_{m_2=1}^{5} y_{m_1} y_{m_2} y'_{m_1} y'_{m_2}$$

$$= y_1 y_1 y'_1 y'_1 + y_1 y_2 y'_1 y'_2 + \cdots + y_1 y_5 y'_1 y'_5$$

$$+ y_2 y_1 y'_2 y'_1 + y_2 y_2 y'_2 y'_2 + \cdots + y_2 y_5 y'_2 y'_5$$

$$+ \cdots$$

$$\vdots$$

$$+ y_5 y_1 y'_5 y'_1 + y_5 y_2 y'_5 y'_2 + \cdots + y_5 y_5 y'_5 y'_5$$

$$= \left(\sum_{m=1}^{5} y_m y'_m \right)^2 = \left(\langle \underline{y}, \underline{y}' \rangle \right)^2$$

SDE exercise p. 170

(1) Show that

$$E\left[\hat{\underline{\Theta}}_{LMMSE}(\underline{Y}) - \underline{\Theta} \right] = \underline{0}$$

and

$$E\left[\left(\hat{\underline{\Theta}}_{LMMSE}(\underline{Y}) - \underline{\Theta} \right) \underline{Y}^T \right] = \underline{\underline{0}}$$

$$\hat{\underline{\Theta}}(\underline{Y}) = \underline{\underline{A}}\, \underline{Y} + \underline{b} \quad \text{with} \quad \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_L \end{bmatrix} \quad \text{and}$$

$$\underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \end{bmatrix}$$

$\lfloor b_K \rfloor$

Usually, $L \geq K$.

L ↓ dimension of measurements $\underline{Y}$

K ↳ dimension of parameter vector $\underline{\theta}$

Note: $\|\underline{x} - \underline{y}\|^2 = \underline{x}^T\underline{x} - 2\underline{x}^T\underline{y} + \underline{y}^T\underline{y}$

Objective function = mean-squared error $\varepsilon(\underline{\underline{A}}, \underline{b})$ between $\hat{\underline{\theta}}_{LMMSE}$ and $\underline{\theta}$

$$\Rightarrow \varepsilon(\underline{\underline{A}}, \underline{b}) = E\left[\|\hat{\underline{\theta}}_{\underline{\underline{A}},\underline{b}}(\underline{Y}) - \underline{\theta}\|^2\right] \text{ ... to be minimized}$$

$$= E\Big[\underbrace{(\underline{\underline{A}}\,\underline{Y} + \underline{b})^T(\underline{\underline{A}}\,\underline{Y} + \underline{b})}_{\|\underline{\underline{A}}\,\underline{Y} + \underline{b}\|^2 = \sum_{i=1}^{K}\left(\sum_{j=1}^{L} a_{ij} Y_j + b_i\right)^2}\Big] - 2E\Big[\underbrace{\underline{\theta}^T(\underline{\underline{A}}\,\underline{Y} + \underline{b})}_{\sum_{i=1}^{K}\theta_i\left(\sum_{j=1}^{L} a_{ij} Y_j + b_i\right)}\Big] + E\left[\|\underline{\theta}\|^2\right]$$

Derivation of $E\left[\hat{\underline{\theta}}_{LMMSE}(\underline{Y}) - \underline{\theta}\right] = \underline{0}$

$$\frac{\partial \varepsilon(\underline{\underline{A}}, \underline{b})}{\partial b_\ell} = E\left[2\sum_{j=1}^{L}(a_{\ell j} Y_j + b_\ell)\right] - 2E\left[\theta_\ell\right] \overset{!}{=} 0$$

$\forall\, \ell = 1, \dots, K$

$$\Rightarrow E\left[(\underline{\underline{A}}\,\underline{Y} + \underline{b})\right] - E\left[\underline{\theta}\right] = \underline{0}$$

$\Leftrightarrow$ $E\left[\hat{\underline{\Theta}}_{LMMSE}(\underline{Y}) - \underline{\Theta}\right] = \underline{0}$

Derivation of $E\left[\left(\hat{\underline{\Theta}}_{LMMSE} - \hat{\underline{\Theta}}\right)\underline{Y}^T\right] = \underline{\underline{0}}$

$$\frac{\partial \varepsilon(\underline{\underline{A}}, \underline{b})}{\partial a_{\ell k}} = E\left[2\underbrace{\left(\sum_{j=1}^{L} a_{\ell j} Y_j + b_\ell\right)}_{\hat{\Theta}_{\ell, LMMSE}} Y_k\right] - 2E\left[\Theta_\ell Y_k\right] \overset{!}{=} 0 \quad \forall\ \ell = 1, \dots, K \text{ and } k = 1, \dots, L$$

$$\Leftrightarrow\ E\left[\underbrace{(\underline{\underline{A}}\,\underline{Y} + \underline{b})}_{\hat{\underline{\Theta}}_{LMMSE}} \underline{Y}^T\right] - E\left[\underline{\Theta}\,\underline{Y}^T\right] = \underline{\underline{0}}$$

$$\Leftrightarrow\ E\left[\left(\hat{\underline{\Theta}}_{LMMSE} - \underline{\Theta}\right)\underline{Y}^T\right] = \underline{\underline{0}}$$

SDE exercise p. 171

2. Consider similarity to a BPSK signal in AWGN

Obs.: $r(t) = b\,p(t) + N(t)$, $\|p(t)\| = 1$, $b \in \{-1, 1\}$

Correlator output:

$R = \langle r(t), p(t) \rangle = B + N$ with $N \sim \mathcal{N}(0; \sigma^2)$

$p_B(b) = \frac{1}{2}\left[\delta(b-1) + \delta(b+1)\right]$ and $\sigma^2 = \frac{N_0}{2}$

$$p_N(n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{n^2}{2\sigma^2} \right\}$$

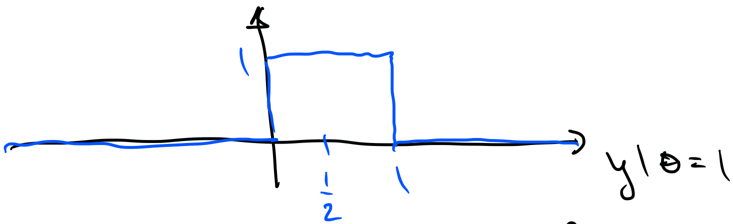
$\Rightarrow$ R non-Gaussian, since

$$p_R(r) = p_B(r) * p_N(r)$$

$$p_\Theta(\theta) = \frac{1}{2}\left\{ \delta(\theta - 1) + \delta(\theta + 1) \right\}$$

Sketch $p(y|\theta=-1) = \mathrm{rect}\left(y + \frac{1}{2}\right) = \mathrm{rect}\left(y - \frac{\theta}{2}\right)$

$p(y|\theta=1) = \mathrm{rect}\left(y - \frac{1}{2}\right) = \mathrm{rect}\left(y - \frac{\theta}{2}\right)$

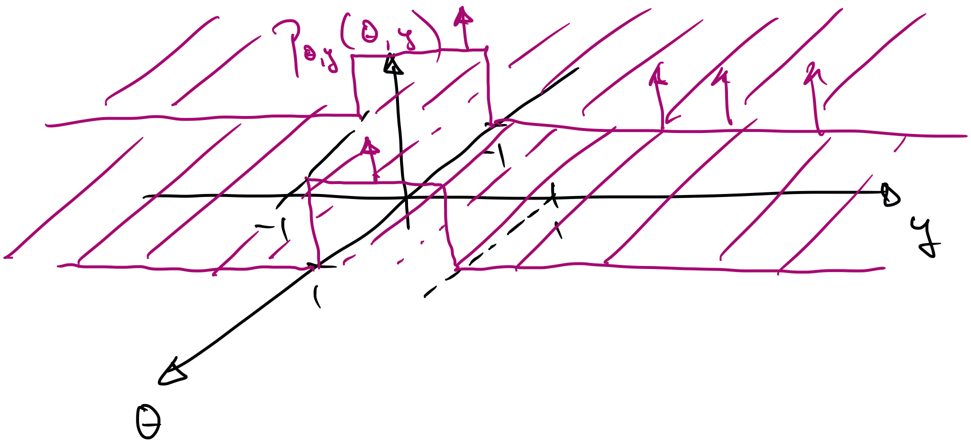


with
$$\mathrm{rect}(x) = \begin{cases} 1 & \text{for } |x| < \frac{1}{2} \\ \frac{1}{2} & \text{" } |x| = \frac{1}{2} \\ 0 & \text{else} \end{cases}$$

$$\Rightarrow p_{\theta,y}(\theta, y) = p_{y|\theta}(y|\theta)\, p_\theta(\theta)$$

$$= \mathrm{rect}\left(y - \frac{\theta}{2}\right) \frac{1}{2}\left\{ \delta(\theta - 1) + \delta(\theta + 1) \right\}$$

(a) Formulate MMSE estimator $\hat{\theta}_{MMSE}(y)$ and quantify MMSE $E\left[(\hat{\theta}_{MMSE}(Y)-\theta)^2\right]$.



$\Rightarrow$ Simply choose $\hat{\theta}(Y) = \text{sign}(Y)$ since this makes the MSE $E\left[(\hat{\theta}(Y)-\theta)^2\right] = 0$

That is, since $\hat{\theta}(Y)$ achieves an MSE of zero, it must be the MMSE estimate of θ, i.e. $\hat{\theta}(y) = \text{sign}(y) = \hat{\theta}_{MMSE}(y)$.