ML Sheet 3

(1) (a) $\mathcal{L}((s(x))_{x\in G_\delta}, \lambda_1, \lambda_2, \lambda_3) = -\sum_{x\in G_\delta} \exp(s(x))\, s(x)\, \delta + \lambda_1 \left(\sum_{x\in G_\delta} \exp(s(x))\, \delta - 1\right)$

$+ \lambda_2 \left(\sum_{x\in G_\delta} \exp(s(x))\, x \delta\right) + \lambda_3 \left(\sum_{x\in G_\delta} \exp(s(x))\, x^2 \delta - \sigma^2\right)$

(b) Let $\hat{x} \in G_\delta$.

$0 \overset{!}{=} \frac{\partial}{\partial s(\hat{x})} \mathcal{L}((s(x))_{x\in G_\delta}, \lambda_1, \lambda_2, \lambda_3)$

$= -\exp(s(\hat{x}))(1 + s(\hat{x}))\delta + \lambda_1 \exp(s(\hat{x}))\delta + \lambda_2 \exp(s(\hat{x}))\hat{x}\delta + \lambda_3 \exp(s(\hat{x}))\hat{x}^2\delta$

$\Leftrightarrow 0 = -(1 + s(\hat{x})) + \lambda_1 + \lambda_2 \hat{x} + \lambda_3 \hat{x}^2$

$\Leftrightarrow s(\hat{x}) = \lambda_3 \hat{x}^2 + \lambda_2 \hat{x} + \lambda_1 - 1 \qquad (\mathrm{I})$

The number λ_3 has to be negative. If this was not the case, then, by equation (I), s would either be constant or tend to $+\infty$ for $\hat{x} \to +\infty$ or for $\hat{x} \to -\infty$. Since G_δ is a regular partition of $\mathbb{R}$, there have to exist sequences in G_δ tending to $\pm\infty$. Now if s is constant or unbounded, this contradicts $0 \overset{!}{=} \frac{\partial \mathcal{L}}{\partial \lambda_1} = \sum_{x\in G_\delta} \exp(s(x))\delta - 1$ (II), the series would diverge to $+\infty$. Hence, $\lambda_3 < 0$.

Rewriting (I) gives

$s(\hat{x}) = \lambda_3 \left(\hat{x}^2 + \frac{\lambda_2}{\lambda_3}\hat{x}\right) + \lambda_1 - 1$

$= \lambda_3 \left(\hat{x} + \frac{1}{2}\cdot\frac{\lambda_2}{\lambda_3}\right)^2 - \frac{1}{4}\cdot\frac{\lambda_2^2}{\lambda_3} + \lambda_1 - 1$

$= -\frac{1}{2\varsigma^2}(\hat{x} - \mu)^2 + r,$

setting $\varsigma^2 := \sqrt{-\frac{1}{2\lambda_3}}$, $\mu := -\frac{1}{2}\cdot\frac{\lambda_2}{\lambda_3}$, $r := -\frac{1}{4}\cdot\frac{\lambda_2^2}{\lambda_3} + \lambda_1 - 1$.

Plugging into equation (II) yields

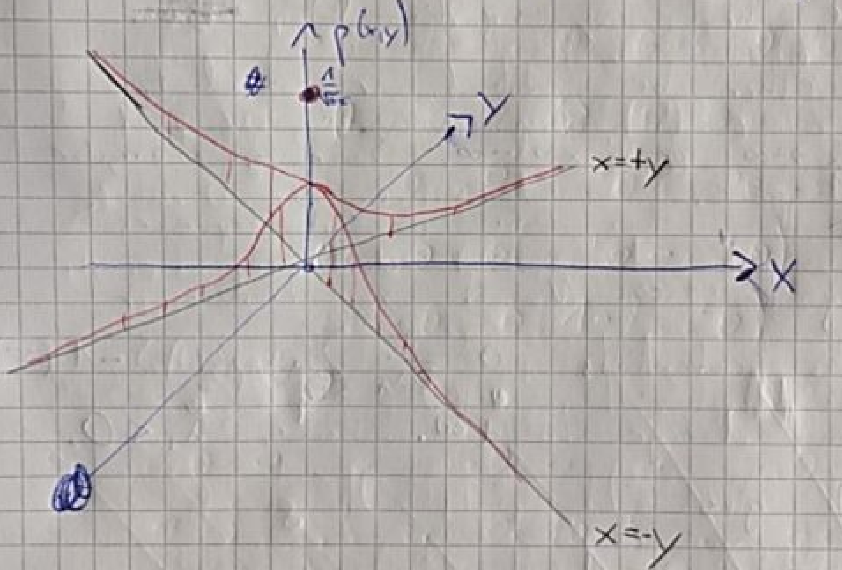
$1 = \sum_{x\in G_\delta} \exp\left(-\frac{1}{2\varsigma^2}(\hat{x} - \mu)^2\right)\cdot \exp(r)\cdot\delta,$

hence $r = -\ln\left(\sum_{x\in G_\delta} \exp\left(-\frac{1}{2\varsigma^2}(\hat{x} - \mu)^2\right)\delta\right).$

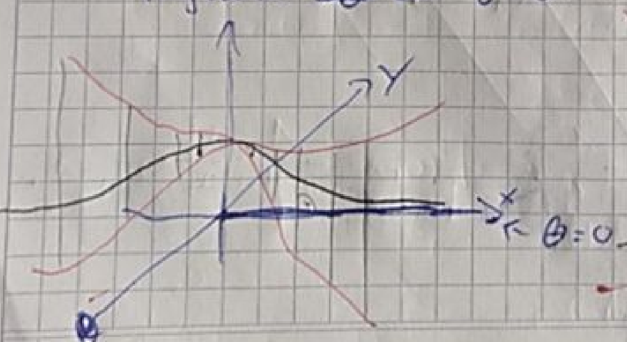
At this point we can already see that ~~p(x) = exp(...)~~ X_3 has a normal distribution with mean μ and variance ϱ^2. By construction, taking the derivatives of $\mathcal{L}$ w.r.t. x_2, x_3 will yield $\mu = 0$, $\varrho^2 = \sigma^2$.

(2) (a) The joint pdf consists of the pdf of a standard normal distribution scaled by the factor $\frac{1}{2}$ along the ~~axis~~ axes $y = x$ and $y = -x$ with a jump to $\frac{1}{\sqrt{2\pi}}$ in 0.

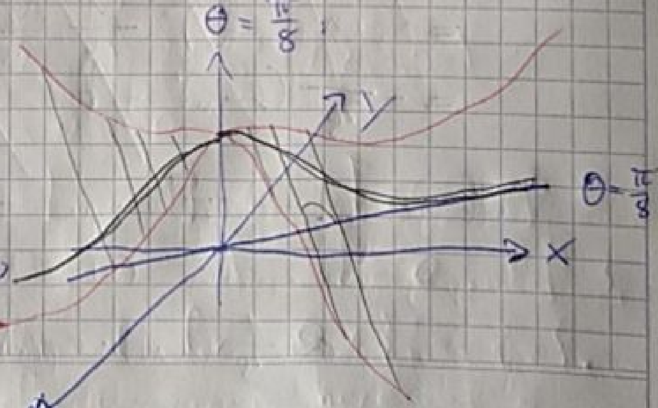
$$p(x,y) = p(x)\,p(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \cdot \begin{cases} 1, & \text{if } x = y = 0 \\ \frac{1}{2}, & \text{if } x = y \neq 0 \text{ or } x = -y \\ 0, & \text{else} \end{cases}$$



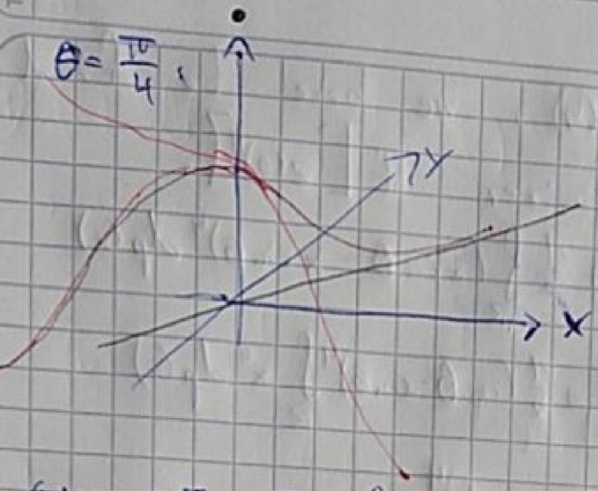
Projection for $\theta = 0$:



$\theta = \frac{\pi}{8}$:



$\theta = \frac{\pi}{4}$,



$\theta = \frac{\pi}{4}$ corresponds to $y=x$

(b) • $\mathbb{E}[z(\theta)] = \int_{\mathbb{R}^2} (x\cos\theta + y\sin\theta)\, dp(x,y)$

$\cancel{\int (x\cos\theta + y\sin\theta)\, dp(x)}$

$= \int_{\mathbb{R}} (x\cos\theta + x\sin\theta)\cdot\frac{1}{2}\, dp(x) + \int_{\mathbb{R}} (x\cos\theta - x\sin\theta)\cdot\frac{1}{2}\, dp(x)$

$= \int_{\mathbb{R}} x\cos\theta\, dp(x)$

$= \cos\theta \quad \mathbb{E}[x]$

$= 0$

• $\mathbb{V}[z(\theta)] = \int_{\mathbb{R}^2} (x\cos\theta + y\sin\theta)^2\, dp(x,y)$

$= \int_{\mathbb{R}} \frac{1}{2}(x\cos\theta + x\sin\theta)^2\, dp(x) + \int_{\mathbb{R}} \frac{1}{2}(x\cos\theta - x\sin\theta)^2\, dp(x)$

$= \int_{\mathbb{R}} \frac{1}{2} x^2 \left((\cos\theta + \sin\theta)^2 + (\cos\theta - \sin\theta)^2\right) dp(x)$

$= \int_{\mathbb{R}} x^2 (\cos^2\theta + \sin^2\theta)\, dp(x)$

$= (\cos^2\theta + \sin^2\theta)\ \mathbb{V}[x]$

$= \cos^2\theta + \sin^2\theta$

$= 1$

Hence, the variance is constant and ~~is~~ maximized by all $\theta \in [0, 2\pi]$

c) By what we have seen in (b), we have to find the value of θ that maximizes $\mathbb{E}[z(\theta)^4]$.

$$\mathbb{E}[z(\theta)^4] = \frac{1}{2}\int_{\mathbb{R}} (x\cos\theta + x\sin\theta)^4 \, dp(x) + \frac{1}{2}\int_{\mathbb{R}} (x\cos\theta - x\sin\theta)^4 \, dp(x)$$

$$= \frac{1}{2}(\cos\theta + \sin\theta)^4 \int_{\mathbb{R}} x^4 \, dp(x) + \frac{1}{2}(\cos\theta - \sin\theta)^4 \int_{\mathbb{R}} x^4 \, dp(x)$$

$$= \frac{3}{2}(\cos\theta + \sin\theta)^4 + \frac{3}{2}(\cos\theta - \sin\theta)^4,$$

where we used that the 4th moment of a standard normal distribution is 3.

$$\frac{d}{d\theta}\mathbb{E}[z(\theta)^4] = 6(\cos\theta + \sin\theta)^3(-\sin\theta + \cos\theta) + 6(\cos\theta - \sin\theta)^3(-\sin\theta - \cos\theta)$$

$$= 6(\cos^3\theta + 3\cos^2\theta\sin\theta + 3\cos\theta\sin^2\theta + \sin^3\theta)(-\sin\theta + \cos\theta)$$

$$- 6(\cos^3\theta - 3\cos^2\theta\sin\theta + 3\cos\theta\sin^2\theta - \sin^3\theta)(\sin\theta + \cos\theta)$$

$$= -12\cos^3\theta\sin\theta + 36\cos^3\theta\sin\theta - 36\cos\theta\sin^3\theta + 12\cos\theta\sin^3\theta$$

$$= 12\cos\theta\sin\theta(-\cos^2\theta + 3\cos^2\theta - 3\sin^2\theta + \sin^2\theta)$$

$$= 24\cos\theta\sin\theta(\cos^2\theta - \sin^2\theta)$$

$$= 24\cos\theta\sin\theta(1 - 2\sin^2\theta)$$

$$\stackrel{!}{=} 0$$

$\Leftrightarrow \quad \cos\theta = 0 \quad$ or $\quad \sin\theta = 0 \quad$ or $\quad \sin\theta = \frac{1}{\sqrt{2}}$

$\Leftrightarrow \quad \theta \in \{\frac{\pi}{2}, \frac{3\pi}{2}\} \quad$ or $\quad \theta \in \{0, \pi\} \quad$ or $\quad \theta \in \{\frac{\pi}{4}, \frac{3\pi}{4}\}$

$\Leftrightarrow \quad \theta \in \{0, \frac{1}{4}\pi, \frac{1}{2}\pi, \frac{3}{4}\pi, \pi, \frac{3}{2}\pi\}$

$$\cancel{\frac{d^2}{d\theta^2}\mathbb{E}[z(\theta)^4] = -24\sin^2\theta(1-2\sin^2\theta) + 24\cos^2\theta(1-2\sin^2\theta) - 96\cos^2\theta\sin^2\theta}$$

Since $\mathbb{E}[z(0)^4] = 3$, $\mathbb{E}[z(\frac{1}{4}\pi)^4] = 6$, $\mathbb{E}[z(\frac{1}{2}\pi)^4] = 3$, $\mathbb{E}[z(\frac{3}{4}\pi)^4] = 6$, $\mathbb{E}[z(\pi)^4] = 3$ and $\mathbb{E}[z(\frac{3}{2}\pi)^4] = 3$, the maximal value is attained for $\theta \in \{\frac{1}{4}\pi, \frac{3}{4}\pi\}$.

I think there's something wrong here, if θ is a solution, then also $\theta + \pi$ should be one.

(3) (a) • $\mathbb{E}[w^T x] = \sum_{k=1}^{d} w_k \underbrace{\mathbb{E}[x_k]}_{=0} = 0$, since $\mathbb{E}[x] = 0 \in \mathbb{R}^d$.

• $\mathbb{V}[w^T x] = \mathbb{E}[(w^T x)^2] = \mathbb{E}[w^T x x^T w] = w^T \underbrace{\mathbb{E}[x x^T]}_{=Id} w = w^T w = \|w\|_2^2 =$

(b) Using (a), the Lagrangian can be written

$$\mathcal{L}(w, \lambda) = \mathbb{E}[(w^T x)^4] + \lambda(\|w\|_2^2 - 1).$$

Supposing we can interchange integration and differentiation, we get

$$0 \doteq \frac{\partial}{\partial w} \mathcal{L}(w, \tilde{\lambda}) = \mathbb{E}[4(w^T x)^3 x] + 2\tilde{\lambda} w$$

$$\Leftrightarrow \underbrace{-\frac{1}{2}\tilde{\lambda}}_{=:\lambda} w = \mathbb{E}[(w^T x)^3 x],$$

~~which is an equivalent problem to the one in the exercise since λ can be recovered by multiplying by w^T from the left and using $\|w\|_2^2 = 1$~~

λ can be found by multiplication of w^T from the left:

$$\lambda \underbrace{w^T w}_{=1} = \mathbb{E}[(w^T x)^3 w^T x] = \mathbb{E}[(w^T x)^4]$$

(c) $F(w) := \mathbb{E}[(w^T x)^3 x] - \lambda w$, i.e., $F(w)_i = \mathbb{E}[(w^T x)^3 x_i] - \lambda w_i$

$\Rightarrow \frac{\partial F(w)_i}{\partial w_j} = \mathbb{E}[3(w^T x)^2 x_i x_j] - \lambda \delta_{ij}$, hence we can write

$$J(w) = \mathbb{E}[3(w^T x)^2 x x^T] - \lambda Id = \mathbb{E}[3(w^T x)^2 x x^T - \lambda Id]$$

Plugging into the ~~formula~~ defining $w^+ = w - J(w)^{-1} F(w)$ yields the formula.

(d)
$$w^+ = w - (\mathbb{E}[3 x x^T (w^T x)^2 - \lambda Id])^{-1} (\mathbb{E}[x (w^T x)^3] - \lambda w)$$

$$= w - (3 \underbrace{\mathbb{E}[x x^T]}_{=Id} \underbrace{\mathbb{E}[(w^T x)^2]}_{=Id, \text{ by (a)}} - \lambda Id)^{-1} (\mathbb{E}[x (w^T x)^3] - \lambda w)$$

$$= w - \frac{1}{3-\lambda} \underbrace{Id^{-1}}_{=Id} (\mathbb{E}[x (w^T x)^3] - \lambda w)$$

$$= -\frac{1}{3-\lambda} \mathbb{E}[x (w^T x)^3] + \underbrace{\left(1 + \frac{\lambda}{3-\lambda}\right)}_{=\frac{3}{3-\lambda}} w = \underbrace{-\frac{1}{3-\lambda}}_{=:\gamma} (\mathbb{E}[x (w^T x)^3] - 3w)$$

2.e) By (b), we have to maximize $\mathbb{E}[z(\theta)]^4$.

- $\mathbb{E}[z(\theta)^4] = \frac{1}{2}\int_{\mathbb{R}} (x\cos\theta + x\sin\theta)^4 \, d\rho(x) + \frac{1}{2}\int_{\mathbb{R}} (x\cos\theta - x\sin\theta)^4 \, d\rho(x)$

$$= \frac{1}{2}\left((\cos\theta + \sin\theta)^4 + (\cos\theta - \sin\theta)^4\right) \underbrace{\int_{\mathbb{R}} x^4 \, d\rho(x)}_{=3}$$

$$= \left(\cos^4\theta + \sin^4\theta + 6\cos^2\theta\sin^2\theta\right) \cdot 3$$

$$= 3\left((\cos^2\theta + \sin^2\theta)^2 + 4\cos^2\theta\sin^2\theta\right)$$

$$= 3\left(1 + \sin^2(2\theta)\right)$$

- $\cancel{\frac{d}{d\theta}\mathbb{E}[z(\theta)^4] = 6\sin(2\theta) \cdot 12 \cdot 6\sin(2\theta)\cdot\cos(2\theta)}$

Since $\sin^2(2\theta) \le 1$, maxima are values of θ so that $\sin(2\theta) = \pm 1$, i.e., $2\theta = \frac{\pi}{2} + k\pi$, $k \in \mathbb{Z}$, hence $\theta \in \{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\}$.