Exercise Sheet 3

1 Maximum Entropy Distribution

a) $$\mathcal{L}((s_k)_{k\in\mathbb{N}}, \lambda_1, \lambda_2) = -\sum_{k=0}^{\infty} e^{s_k} \cdot s_k + \lambda_1\left(\sum_{k=0}^{\infty} k e^{s_k} - 1\right) + \lambda_2\left(\sum_{k=0}^{\infty} e^{s_k} - 1\right)$$

$$= \sum_{k=0}^{\infty} e^{s_k}(-s_k) + \sum_{k=0}^{\infty} e^{s_k}(\lambda_1 k) - \lambda_1 + \sum_{k=0}^{\infty} e^{s_k}(\lambda_2) - \lambda_2$$

$$= \sum_{k=0}^{\infty} e^{s_k}(\lambda_1 k + \lambda_2 - s_k) - \lambda_1 - \lambda_2$$

b) $$\frac{\partial}{\partial s_k}\mathcal{L}(\cdot) \overset{!}{=} 0 \quad \forall k \in \mathbb{N}$$

$$\frac{\partial}{\partial s_k}\mathcal{L}((s_k)_{k\in\mathbb{N}}, \lambda_1, \lambda_2) = e^{s_k}(\lambda_1 k + \lambda_2 - s_k) - e^{s_k} = 0$$

$$\Leftrightarrow e^{s_k}(\lambda_1 k + \lambda_2 - s_k - 1) = 0 \Leftrightarrow s_k = \lambda_1 k + \lambda_2 - 1$$

$$Pr(X=k) = e^{\lambda_1 k + \lambda_2 - 1} = e^{\lambda_2 - 1} e^{\lambda_1 k}$$

$$= \underbrace{e^{\lambda_2 - 1}}_{\alpha} \underbrace{e^{\lambda_1}}_{\beta}{}^{k} \Rightarrow \alpha\beta^k$$

c) I $$\frac{\partial \mathcal{L}(\cdot)}{\partial s_k} \overset{!}{=} 0 \Leftrightarrow P(X=k) = \alpha\beta^k$$

II $$\frac{\partial \mathcal{L}(\cdot)}{\partial \lambda_2} \overset{!}{=} 0 \Leftrightarrow \sum_{k=0}^{\infty} P(X=k) = 1 \overset{(I)}{\Leftrightarrow} \sum_{k=0}^{\infty} \alpha\beta^k = 1$$

$$\Leftrightarrow \frac{\alpha}{1-\beta} = 1 \Leftrightarrow \alpha = 1-\beta$$

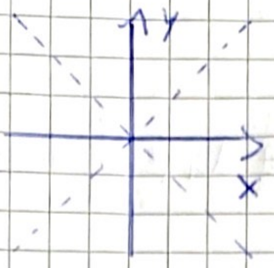
III $$\frac{\partial \mathcal{L}(\cdot)}{\partial \lambda_1} \overset{!}{=} 0 \Leftrightarrow \sum_{k=0}^{\infty} k P(X=k) = 1 \overset{(I)}{\Leftrightarrow} \sum_{k=0}^{\infty} k\alpha\beta^k = 1$$

$$\Leftrightarrow \alpha\sum_{k=0}^{\infty} k\beta^k = 1 \Leftrightarrow \alpha\frac{\beta}{(1-\beta)^2} = 1 \overset{(II)}{\Leftrightarrow} \frac{(1-\beta)\beta}{(1-\beta)^2} = 1$$

$$\Leftrightarrow \beta = (1-\beta) \Leftrightarrow \beta = 0{,}5$$

2. Independent Components in Two Dimensions

a)



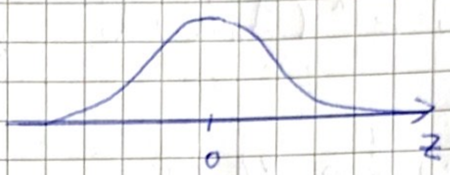
$p(x) \rightarrow$ sample over x according to Gaussian dist.

$p(y|x) \rightarrow$ "p of y given x"

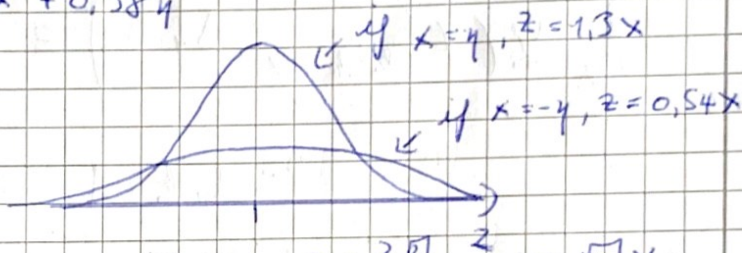
$$p(y|x) = \frac{1}{2}\,\underbrace{\delta(y-x)}_{y=x} + \frac{1}{2}\,\underbrace{\delta(y+x)}_{y=-x}$$

$\downarrow$ in 50% of samples

If $\vartheta = 0$: $z(\vartheta) = x$



If $\vartheta = \frac{1}{8}\pi$: $z = 0{,}92x + 0{,}38y$

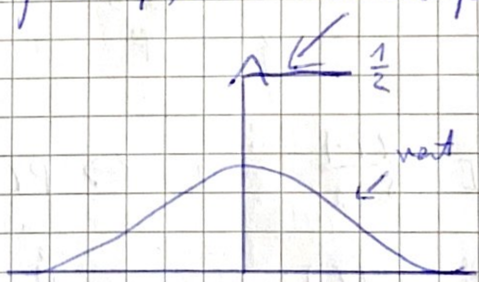


If $\vartheta = \frac{\pi}{4}\pi$: $z = \frac{\sqrt{2}}{2}(x+y)$

$\rightarrow$ if $x = y$, $z = \frac{2\sqrt{2}}{2}x = \sqrt{2}\,x$

$\rightarrow$ Gaussian is "lower" $\rightarrow$ max variance

$\rightarrow$ if $x = -y$, $z = 0$ $\rightarrow$ dirac peak



b) $E[z] = E[x\cos\vartheta + y\sin\vartheta] = \cos\vartheta \underbrace{E[x]}_{=0} + \sin\vartheta\, E[y]$

$= \sin\vartheta \, \underset{x}{E}[\underset{y|x}{E}[y]] = \sin\vartheta \, \underset{x}{E}[0] = 0$

$Var[z] = E[(z - \underbrace{E[z]}_{=0})^2] = E[z^2]$

$= E[x^2\cos^2\vartheta + xy\cos\vartheta\sin\vartheta + y^2\sin\vartheta]$

$= \cos^2\vartheta \underbrace{E[x^2]}_{=1} + \cos\vartheta\sin\vartheta \underbrace{E[xy]}_{=0} + \sin\vartheta \underbrace{E[y^2]}_{=1}$

$= \cos^2\vartheta + \sin^2\vartheta = 1$

c) $E[z(\vartheta)] = 0$, $Var[z(\vartheta)] = 1$

$kurt[z(\vartheta)] = \frac{E[(z(\vartheta) - 0)^4]}{1^2} - 3 = E[z(\vartheta)^4] - 3$

$= E[(x\cos\vartheta + y\sin\vartheta)^4] - 3$

$= E[(x^2\cos^2\vartheta + 2xy\cos\vartheta\sin\vartheta + y^2\sin^3\vartheta)^2] - 3$

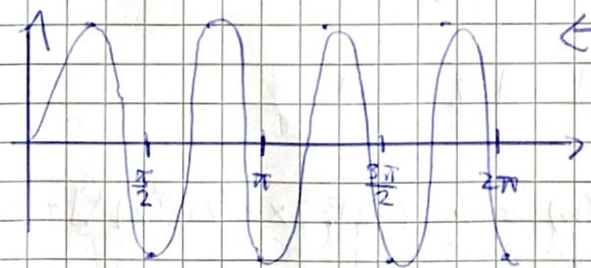
$= \cos^4\vartheta\, E[x^4] + 4\cos^2\vartheta\sin\vartheta\, E[x^3y] + 6\cos^2\vartheta\sin^2\vartheta\, E[x^2y^2]$

$+ 4\cos\vartheta\sin^3\vartheta\, E[xy^3] + \sin^4\vartheta\, E[y^4] - 3$

$= E[x^4]((\underbrace{\cos^2\vartheta + \sin^2\vartheta}_{=1}) + 4\cos^2\vartheta\sin^2\vartheta) - 3$

$= E[x^4](1 + 1 - \frac{1 - \cos 4\vartheta}{2}) - 3$ → $\cos 4\vartheta$ needs to be maximized!

← max.



→ Independent components:

$\frac{\pi}{4}$, $\frac{3\pi}{4}$, $\frac{5\pi}{4}$, $\frac{7\pi}{4}$

3 Deriving a Special Case of Fast ICA

a) $$\underbrace{\mathbb{E}_x[w^T x]}_{\text{mean of } w^T x} = w^T \underbrace{\mathbb{E}_x[x]}_{=0} = 0$$

$$Var[w^T x] = \mathbb{E}_x[(w^T x)^2] = \mathbb{E}_x[(w^T x w^T x)]\ \mathbb{E}[w^T x x^T w]$$

$$= w^T \underbrace{\mathbb{E}_x[x x^T]}_{=1} w = w^T w = \|w\|^2 = 1$$

b) $$kurt[w^T x] = \frac{\mathbb{E}[(w^T x - \overbrace{\mathbb{E}[w^T x]}^{=0})^4]}{\underbrace{(Var[w^T x])^2}_{=1}} - 3$$

$$= \mathbb{E}[(w^T x)^4] - 3$$

$$\mathcal{L}(w, \lambda) \Rightarrow \mathbb{E}[(w^T x)^4] - 3 + \lambda 2 (1 - \|w\|) = 0$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial w} \Rightarrow \mathbb{E}[\tfrac{\partial}{\partial w}(w^T x)^4] + \lambda 2 (-\tfrac{\partial}{\partial w} w^T w) = 0$$

$$\Leftrightarrow \mathbb{E}[4 (w^T x)^3 x^T] - 4 \lambda w^T = 0$$

$$\Leftrightarrow \mathbb{E}[(w^T x)^3 x] - \lambda w$$

$$\Leftrightarrow \lambda w = \mathbb{E}[x (w^T x)^3]$$

c) $$F(w) = \mathbb{E}[x (w^T x)^3] - \lambda w = 0$$

$$J(w) = \frac{\partial F(w)}{\partial w} = \mathbb{E}[\tfrac{\partial}{\partial w} x (w^T x)^3] - \lambda = 0$$

$$= \mathbb{E}[3 x x^T (w^T x)^2] - \lambda \mathbb{1} = 0$$

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

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

$J(w) = E[3xx^T(w^Tx)^2] = 3E[xx^T]\,E[(w^Tx)^2] - \lambda\mathbb{1}$

$= 3\mathbb{1} - \lambda\mathbb{1} = \mathbb{1}(3-\lambda)$

$w^+ = w - \mathbb{1}^{-1}(3-\lambda)\left(E[x(w^Tx)^3] - \lambda w\right)$

$\Leftrightarrow w^+ = w - \dfrac{E[x(w^Tx)^3] - \lambda w}{3-\lambda}$

$\Leftrightarrow (3-\lambda)\,w^+ = 3w - \lambda w - E[x(w^Tx)^3] + \lambda w$

$\Leftrightarrow \cancel{4}(3-\lambda)\,w^+ = -E[x(w^Tx)^3] + 3w$

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