Q1 a) $\mu_y = \int y\, p(y)\, dy \overset{\text{sum rule}}{=} \iint y\, p(y,x)\, dy\, dx \overset{\text{product rule}}{=} \iint y\, p(y|x)\, p(x)\, dx$

$= \int \mu_{y|x}(x)\, p(x)\, dx = \mathbb{E}_{p(x)}\left[\mu_{y|x}(x)\right]$

b) $\mu_d = \mathbb{E}_{p(v)}\left[cv^2\right] = c\left(\mu_v^2 + \sigma_v^2\right)$

Q2

a) $p(x|a,v) = \dfrac{p(x)\, p(a|x)\, p(v|x)}{p(a,v)}$

b) $p(x|a,v) \propto N(x; 0, \sigma_0^2) \cdot N(a; x, \sigma_a^2) \cdot N(v; x, \sigma_v^2) \propto N(x; \mu_{x|a,v}, \sigma^2_{x|a,v})$

$\mu_{x|a,v} = \sigma^2_{x|a,v}\left(\dfrac{0}{\sigma_0^2} + \dfrac{a}{\sigma_a^2} + \dfrac{v}{\sigma_v^2}\right) = \sigma^2_{x|a,v}\left(\dfrac{a}{\sigma_a^2} + \dfrac{v}{\sigma_v^2}\right)$

$\sigma^2_{x|a,v} = \left(\dfrac{1}{\sigma_0^2} + \dfrac{1}{\sigma_a^2} + \dfrac{1}{\sigma_v^2}\right)^{-1}$

c) $\sigma_a^2 \to \infty \Rightarrow \sigma^2_{x|a,v} \to \left(\dfrac{1}{\sigma_0^2} + \dfrac{1}{\sigma_v^2}\right)^{-1} = \dfrac{\sigma_v^2 \sigma_0^2}{\sigma_v^2 + \sigma_0^2}$

$\mu_{x|a,v} \to \dfrac{\sigma_0^2}{\sigma_v^2 + \sigma_0^2}\, v$

just like the audio signal was not observed i.e. same posterior as $p(x|v)$

c) An off-the-shelf implementation of the Kalman filter uses first order Markov dynamics for the hidden state so is not immediately applicable to the first form of the model. The second form of the model is in the standard form & so the Kalman filter can now be used for inference.

Q3

a) $\mathcal{L}(m) = \log \prod_n p(y_n | x_n, m, c, \alpha) = -\frac{1}{2} \sum_n \log(\alpha x_n^2) + \frac{(m x_n + c - y_n)^2}{\alpha x_n^2}$

b) $$\frac{d\mathcal{L}(m)}{dm} = -\sum_n \frac{1}{\alpha x_n^2} x_n (m x_n + c - y_n) = 0$$

$$\Rightarrow \quad Nm + \sum_n \frac{(c - y_n)}{x_n} = 0$$

$$\therefore \quad m = \frac{1}{N} \sum_n \frac{(y_n - c)}{x_n}$$

Two effects

c) - Select smallest x_n value possible ($x_n = 1$) since noise is smallest there

- select largest x_n since then signal is largest as $y_n = \underbrace{m x_n + c}_{\text{signal}} + \underbrace{\epsilon_n}_{\text{noise}}$

Here these two effects cancel (signal to noise constant across space)

Lots of ways to see this eg

sub in for this

$n_n \sim N(0,1)$

estimator for m $\rightarrow$ $$m_{MLE} = \frac{1}{N} \sum_n \frac{(y_n - c)}{x_n} = \frac{1}{N} \sum_n \frac{(m x_n + c + \alpha x_n n_n - c)}{x_n}$$

true value

true slope

$$m_{MLE} = \frac{1}{N} \sum_n \underbrace{(m + \alpha n_n)}$$

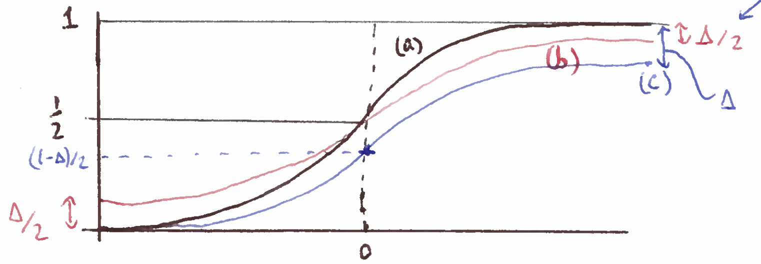
noise with fixed variance

terms in this sum statistically identical

$\Rightarrow$ doesn't matter how we set x_n

Q4.

a) $p(y=1|w,x) = \sigma(w^Tx) = \dfrac{1}{1+e^{-w^Tx}}$



even though we lapse with probability Δ in (b) we still get half correct when this happens

b) l

$p(l=0) = 1-\Delta$ $\quad$ $p(l=1) = \Delta$

$p(y=1|w,x,l=0) = \sigma(w^Tx)$ $\quad$ $p(y=1|l=1) = \frac{1}{2}$

$$p(y=1|w,x) = \sum_l p(y=1|w,x,l)\,p(l) = (1-\Delta)\sigma(w^Tx) + \Delta/2$$

c) l

$p(l=0) = 1-\Delta$

$p(y=1|w,x,l=0) = \sigma(w^Tx)$ $\quad$ $p(y=1|l=1) = 0$

$$p(y=1|w,x) = \sum_l p(y=1|w,x,l)\,p(l) = (1-\Delta)\sigma(w^Tx)$$

like (b) but shifted down by $\Delta/2$

Q5

a) MoG model

$$p(z_n = k) = \tilde{\pi}_k \quad \text{for } k = 1 \ldots K$$

mixing proportions

number of components

$$p(\underline{y}_n \mid z_n = k) = \mathcal{N}(\underline{y}_n ; \underline{\mu}_k, \underline{\underline{\Sigma}}_k)$$

component means

component covariances

$$\Rightarrow p(\underline{y}_n \mid \theta) = \sum_{k=1}^{K} \tilde{\pi}_k \, \mathcal{N}(\underline{y}_n ; \underline{\mu}_k, \underline{\underline{\Sigma}}_k)$$

b)

	original model			MoG	
state	prior	conditional	state	prior	conditional
	$p(s_1, s_2)$	$p(\underline{y} \mid s_1, s_2, \theta)$		$p(z)$	$p(\underline{y} \mid z, \tilde{\theta})$
$s_1 = 0, s_2 = 0$	$(1-\pi_1)(1-\pi_2)$	$\mathcal{N}(\underline{y} ; \underline{0}, \underline{\underline{I}}\sigma_y^2)$	$z = 1$	$\tilde{\pi}_1$	$\mathcal{N}(\underline{y} ; \underline{0}, \underline{\underline{I}}\sigma_y^2)$
$s_1 = 1, s_2 = 0$	$\pi_1(1-\pi_2)$	$\mathcal{N}(\underline{y} ; \underline{w}_1, \underline{\underline{I}}\sigma_y^2)$	$z = 2$	$\tilde{\pi}_2$	$\mathcal{N}(\underline{y} ; \underline{w}_1, \underline{\underline{I}}\sigma_y^2)$
$s_1 = 0, s_2 = 1$	$(1-\pi_1)\pi_2$	$\mathcal{N}(\underline{y} ; \underline{w}_2, \underline{\underline{I}}\sigma_y^2)$	$z = 3$	$\tilde{\pi}_3$	$\mathcal{N}(\underline{y} ; \underline{w}_2, \underline{\underline{I}}\sigma_y^2)$
$s_1 = 1, s_2 = 1$	$\pi_1 \pi_2$	$\mathcal{N}(\underline{y} ; \underline{w}_1 + \underline{w}_2, \underline{\underline{I}}\sigma_y^2)$	$z = 4$	$\tilde{\pi}_4$	$\mathcal{N}(\underline{y} ; \underline{w}_1 + \underline{w}_2, \underline{\underline{I}}\sigma_y^2)$

equal to

μ_1 $\underline{\underline{\Sigma}}_1$ μ_2 $\underline{\underline{\Sigma}}_2$ μ_3 $\underline{\underline{\Sigma}}_3$ μ_4 $\underline{\underline{\Sigma}}_4$

Q6

a) let $p(s_{1,n}=0,\ s_{2,n}=0 \mid \underline{y}_n) = \dfrac{p(\underline{y}_n \mid s_{1,n}=0,\ s_{2,n}=0)\ p(s_{1,n}=0,\ s_{2,n}=0)}{\sum_{s_1 s_2} p(\underline{y}_n \mid s_{1,n}, s_{2,n})\ p(s_{1,n}, s_{2,n})} = \dfrac{r_{00}^{(n)}}{r^{(n)}}$

similarly

$$p(s_{1n}=1,\ s_{2n}=0 \mid \underline{y}_n) = \frac{r_{10}^{(n)}}{r^{(n)}}$$

$$p(s_{1n}=0;\ s_{2n}=1 \mid \underline{y}_n) = \frac{r_{01}^{(n)}}{r^{(n)}}$$

$$p(s_{1n}=1,\ s_{2n}=1 \mid \underline{y}_n) = \frac{r_{11}^{(n)}}{r^{(n)}}$$

where $r^{(n)} = r_{00}^{(n)} + r_{10}^{(n)} + r_{01}^{(n)} + r_{11}^{(n)}$

and $r_{00}^{(n)} = (1-\pi_1)(1-\pi_2)\ \mathcal{N}(\underline{y}_n; \underline{0}, \underline{\underline{I}}\sigma_y^2)$

$r_{10}^{(n)} = \pi_1(1-\pi_2) \cdot \mathcal{N}(\underline{y}_n; \underline{w}_1, \underline{\underline{I}}\sigma_y^2)$

$r_{01}^{(n)} = (1-\pi_1)\pi_2 \cdot \mathcal{N}(\underline{y}_n; \underline{w}_2, \underline{\underline{I}}\sigma_y^2)$

$r_{11}^{(n)} = \pi_1\pi_2 \cdot \mathcal{N}(\underline{y}_n; \underline{w}_1+\underline{w}_2, \underline{\underline{I}}\sigma_y^2)$

b) M step computes $\underset{\theta}{\arg\max} \sum_n \mathbb{E}_{q_n}\left[\log p(s_{1n} s_{2n} \mid \theta) + \log p(\underline{y}_n \mid s_{1n}, s_{2n}, \theta)\right]$

- differentiate and
- either

$\pi_1 \& \pi_2$ eg $\rho_y = \log\sigma_y^2$

set to zero to find optimum (using Lagrange multiplier / reparameterisation appropriately)

use a gradient based update

Q7

a)

$$x_t = \lambda x_{t-1} + \sqrt{1-\lambda^2}\, \varepsilon_t \qquad \varepsilon_t \sim N(0,1)$$

$\lambda = 0.99$

$\Rightarrow$ strong autocorrelation close to 1

$\sigma_x \Rightarrow$ ensures marginal variance of x_t is 1

b) $$y_t \mid x_t \sim N\left(0, \left[\log\left(1 + e^{2x_t}\right)\right]^2\right)$$

y is zero mean

variance depends on x_t

$\Rightarrow$ higher x_t $\Rightarrow$ larger variance

more negative $x_t \Rightarrow$ lower variance.

c) useful for modelling

i) volatility of stocks & shares

ii) natural sounds c.f. related to amplitude modulation

Q8

a) The standard linear Gaussian state space model has a first order markov hidden state

$$x_t = \lambda x_{t-1} + \sigma_x \varepsilon_t$$

The model here has a second order markov hidden state

$$x_t = \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \sigma_x \varepsilon_t$$

b) $$\begin{bmatrix} a, & b \\ c, & d \end{bmatrix} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} \sigma_x & 0 \\ 0 & 0 \end{bmatrix}$$

$\Rightarrow$ $z_{1t} = x_t = \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \sigma_x \varepsilon_{1t}$

$z_{2t} = x_{t-1}$

$[i, j] = [1, 0] \Rightarrow y_t = x_t + \eta_t \sigma_y$