Philipp von Bachmann
Laura Häge

| 1 | 2 | Σ |
|---|---|---|
| | | |

Assignment Nr. 2

# 1 Maximum Likelihood and Maximum A Posteriori Estimation

## (a)

$$\begin{aligned} w_{ML} &= argmax_w p(y|x,w) \\ &= argmax_w log(\frac{1}{\sqrt{2\pi g(x)}} exp(-\frac{(y-\langle w,x\rangle)^2}{2g(x)})) \\ &= argmax_w log(exp(-\frac{(y-\langle w,x\rangle)^2}{2g(x)})) \\ &= argmax_w log(exp(-(y-\langle w,x\rangle)^2)) \\ &= argmax_w -(y-\langle w,x\rangle)^2 \\ &= argmin_w (y-\langle w,x\rangle)^2 \end{aligned}$$

Compute derivative:

$$\nabla_w (y-\langle w,x\rangle)^2 = -2yx + 2\langle w,x\rangle x$$

Set to 0:

$$\begin{aligned} 0 &= -2yx + 2\langle w,x\rangle x \\ 2yx &= 2\langle w,x\rangle x \\ yx &= \langle w,x\rangle x \\ y &= \langle w,x\rangle \\ w &= yx^{-1} \end{aligned}$$

## (b)

$$\begin{aligned} p(x,y|w) &= \frac{p(x,y,w)}{p(w)} \\ &= \frac{p(y|x,w)p(x|w)p(w)}{p(w)} \\ &= p(y|x,w)p(x|w) \\ &= p(y|x,w)p(x) \end{aligned}$$

$$
\begin{aligned}
w_{MAP} &= argmax_w \sum_i log(p(y_i|x_i, w)) + log(p(w)) \\
&= argmax_w \sum_i log(\frac{1}{\sqrt{2\pi g(x_i)}} exp(-\frac{(y_i - \langle w, x_i\rangle)^2}{2g(x_i)})) + log(\frac{1}{(2\pi)^{\frac{d}{2}}(\prod_j \lambda_j)^{\frac{1}{2}}} exp(-\langle w, \Lambda^{-1}w\rangle)) \\
&= argmax_w \sum_i log(exp(-(y_i - \langle w, x_i\rangle)^2)) + log(exp(-\langle w, \Lambda^{-1}w\rangle) \\
&= argmax_w - \sum_i (y_i - \langle w, x_i\rangle)^2 - \langle w, \Lambda^{-1}w\rangle \\
&= argmax_w - \sum_i (y_i^2 - 2y_i\langle w, x_i\rangle + \langle w, x_i\rangle^2) - \langle w, \Lambda^{-1}w\rangle \\
&= argmin_w \sum_i y_i^2 - 2y_i\langle w, x_i\rangle + \langle w, x_i\rangle^2 + \langle w, \Lambda^{-1}w\rangle
\end{aligned}
$$

Compute derivative:

$$
\begin{aligned}
&\nabla_w \sum_i y_i^2 - 2y_i\langle w, x_i\rangle + \langle w, x_i\rangle^2 + \langle w, \Lambda^{-1}w\rangle \\
&= \sum_i -2y_i x_i + 2\langle w, x_i\rangle x_i + 2\langle w, \Lambda^{-1}\rangle
\end{aligned}
$$

Set to 0:

$$
\begin{aligned}
0 &= \sum_i -2y_i x_i + 2\langle w, x_i\rangle x_i + 2\langle w, \Lambda^{-1}\rangle \\
0 &= \sum_i -y_i x_i + \langle w, x_i\rangle x_i + \langle w, \Lambda^{-1}\rangle \\
\sum_i y_i x_i &= \sum_i \langle w, x_i\rangle x_i + \langle w, \Lambda^{-1}\rangle \\
\sum_i y_i &= \sum_i \langle w, x_i\rangle + \langle w, \Lambda^{-1}x_i^{-1}\rangle \\
\sum_i y_i &= \sum_i \langle w, x_i + \Lambda^{-1}x_i^{-1}\rangle \\
\sum_i \frac{y_i}{x_i^{-1} + x_i\Lambda^{-1}} &= w
\end{aligned}
$$

## ML and MAP estimators

### (a)

Formulate the problem and simplify:

$$\begin{aligned}
\theta_{ML} &= argmax_\theta log(p(a_1, \ldots, a_n, b_1, \ldots, b_n|\theta)) \\
&= argmax_\theta \sum_i log(p(a_i|\theta))) + \sum_i log(p(b_i|\theta))) \\
&= argmax_\theta \sum_i log(\frac{1}{\sqrt{2\pi\sigma_1^2}} exp(-\frac{(a_i-\theta)^2}{2\sigma_1^2})) + \sum_i log(\frac{1}{\sqrt{2\pi\sigma_2^2}} exp(-\frac{(b_i-\theta)^2}{2\sigma_2^2})) \\
&= argmax_\theta \sum_i -\frac{(a_i-\theta)^2}{2\sigma_1^2} + \sum_i -\frac{(b_i-\theta)^2}{2\sigma_2^2} \\
&= argmin_\theta \sum_i \frac{(a_i-\theta)^2}{2\sigma_1^2} + \sum_i \frac{(b_i-\theta)^2}{2\sigma_2^2}
\end{aligned}$$

Compute derivative:

$$\begin{aligned}
&\frac{\partial}{\partial\theta} \sum_i \frac{(a_i-\theta)^2}{2\sigma_1^2} + \sum_i \frac{(b_i-\theta)^2}{2\sigma_2^2} \\
&= \frac{\partial}{\partial\theta} \sum_i \frac{a_i^2 - 2a_i\theta - \theta^2}{2\sigma_1^2} + \sum_i \frac{b_i^2 - 2b_i\theta + \theta^2}{2\sigma_2^2} \\
&= \sum_i \frac{-2a_i + 2\theta}{2\sigma_1^2} + \sum_i \frac{-2b_i + 2\theta}{2\sigma_2^2} \\
&= \sum_i \frac{-a_i + \theta}{\sigma_1^2} + \sum_i \frac{-b_i + \theta}{2\sigma_2^2} \\
&= \frac{n\theta}{\sigma_1^2} + \sum_i -\frac{a_i}{\sigma_1^2} + \frac{n\theta}{\sigma_2^2} + \sum_i -\frac{a_i}{\sigma_2^2}
\end{aligned}$$

Set to 0:

$$\begin{aligned}
0 &= \frac{n\theta}{\sigma_1^2} + \sum_i -\frac{a_i}{\sigma_1^2} + \frac{n\theta}{\sigma_2^2} + \sum_i -\frac{b_i}{\sigma_2^2} \\
0 &= n\theta\sigma_2^2 + \sum_i -\sigma_2^2 a_i + n\theta\sigma_1^2 + \sum_i -\sigma_1^2 b_i \\
\sum_i \sigma_2^2 a_i + \sum_i \sigma_1^2 b_i &= n\theta\sigma_2^2 + n\theta\sigma_1^2 \\
\sum_i \sigma_2^2 a_i + \sum_i \sigma_1^2 b_i &= (\sigma_2^2 + \sigma_1^2) n\theta \\
\theta &= \frac{\sum_i \sigma_2^2 a_i + \sum_i \sigma_1^2 b_i}{n(\sigma_2^2 + \sigma_1^2)}
\end{aligned}$$

## (b)

$$
\begin{aligned}
\theta_{MAP} &= argmax_\theta log(p(a_1,\dots,a_n,b_1,\dots,b_n|\theta)) + log(p(\theta)) \\
&= argmax_\theta \sum_i log(p(a_i|\theta))) + \sum_i log(p(b_i|\theta))) + log(p(\theta)) \\
&= argmax_\theta \sum_i log(\frac{1}{\sqrt{2\pi\sigma_1^2}} exp(-\frac{(a_i-\theta)^2}{2\sigma_1^2})) + \sum_i log(\frac{1}{\sqrt{2\pi\sigma_2^2}} exp(-\frac{(b_i-\theta)^2}{2\sigma_2^2})) \\
&+ log(\frac{1}{\sqrt{2\pi\sigma_P^2}} exp(-\frac{(\mu_P-\theta)^2}{2\sigma_P^2})) \\
&= argmax_\theta \sum_i -\frac{(a_i-\theta)^2}{2\sigma_1^2} + \sum_i -\frac{(b_i-\theta)^2}{2\sigma_2^2} - \frac{(\mu_P-\theta)^2}{2\sigma_P^2}
\end{aligned}
$$

Compute derivative:

$$
\begin{aligned}
&\frac{\partial}{\partial\theta} \sum_i \frac{(a_i-\theta)^2}{2\sigma_1^2} + \sum_i \frac{(b_i-\theta)^2}{2\sigma_2^2} - \frac{(\mu_P-\theta)^2}{2\sigma_P^2} \\
&= \frac{\partial}{\partial\theta} \sum_i \frac{a_i^2 - 2a_i\theta - \theta^2}{2\sigma_1^2} + \sum_i \frac{b_i^2 - 2b_i\theta + \theta^2}{2\sigma_2^2} - \frac{\mu_P^2 - 2\mu_P\theta + \theta^2}{2\sigma_P^2} \\
&= \sum_i \frac{-2a_i + 2\theta}{2\sigma_1^2} + \sum_i \frac{-2b_i + 2\theta}{2\sigma_2^2} - \frac{-2\mu_P + 2\theta}{2\sigma_P^2} \\
&= \sum_i \frac{-a_i + \theta}{\sigma_1^2} + \sum_i \frac{-b_i + \theta}{2\sigma_2^2} + \frac{\mu_P - \theta}{\sigma_P^2} \\
&= \frac{n\theta}{\sigma_1^2} + \sum_i -\frac{a_i}{\sigma_1^2} + \frac{n\theta}{\sigma_2^2} + \sum_i -\frac{a_i}{\sigma_2^2} + \frac{\mu_P}{\sigma_P^2} - \frac{\theta}{\sigma_P^2}
\end{aligned}
$$

Set to 0:

$$
\begin{aligned}
0 &= \frac{n\theta}{\sigma_1^2} + \sum_i -\frac{b_i}{\sigma_1^2} + \frac{n\theta}{\sigma_2^2} + \sum_i -\frac{a_i}{\sigma_2^2} + \frac{\mu_P}{\sigma_P^2} - \frac{\theta}{\sigma_P^2} \\
\sum_i \sigma_2^2\sigma_P^2 a_i + \sum_i \sigma_P^2\sigma_1^2 b_i - \sigma_1^2\sigma_2^2\mu_P &= \sigma_2^2\sigma_P^2 n\theta + \sigma_1^2\sigma_P^2 n\theta - \theta\sigma_1^2\sigma_2^2 \\
\sum_i \sigma_2^2\sigma_P^2 a_i + \sum_i \sigma_P^2\sigma_1^2 b_i - \sigma_1^2\sigma_2^2\mu_P &= (\sigma_2^2 + \sigma_1^2) n\sigma_P^2\theta - \theta\sigma_1^2\sigma_2^2 \\
\sum_i \sigma_2^2\sigma_P^2 a_i + \sum_i \sigma_P^2\sigma_1^2 b_i - \sigma_1^2\sigma_2^2\mu_P &= ((\sigma_2^2 + \sigma_1^2) n\sigma_P^2 - \sigma_1^2\sigma_2^2)\theta \\
\frac{\sum_i \sigma_2^2\sigma_P^2 a_i + \sum_i \sigma_P^2\sigma_1^2 b_i - \sigma_1^2\sigma_2^2\mu_P}{((\sigma_2^2 + \sigma_1^2) n\sigma_P^2 - \sigma_1^2\sigma_2^2)} &= \theta
\end{aligned}
$$

We would get $\hat{\theta}_{ML} = \hat{\theta}_{MAP}$ if $\mu_p = 0, \sigma_P^2 = 1$.