Philipp Lintl
12152498
lintl.philipp@gmail.com

Homework Assignment 2
Machine Learning 1, 18/19

2018-11-15

1 MAP solution with correlated responses

Given

- *correlated measurements*: known, nonsingular covariance matrix $\mathbf{Omega} \rightarrow$ no iid samples anymore
- N training vectors $\{\mathbf{x_1}, \dots, \mathbf{x_N}\}$
- with feature vector mapping $\boldsymbol{\phi}_i = (\phi_0(\mathbf{x_i}), \phi_1(\mathbf{x_i}), \dots, \phi_{M-1}(\mathbf{x_i})^T$
- with basis functions: $\phi_j(\mathbf{x})$
- training set tuples $(\mathbf{x_n}, t_n)$
- regression prediction: $y(\mathbf{x_n}), \mathbf{w} = \mathbf{w}^T \phi_\mathbf{n}$
- *whole dataset likelihood* is Gaussian: $p(\mathbf{t}|\boldsymbol{\Phi}|\boldsymbol{w}, \boldsymbol{\Omega}) = \mathcal{N}(\mathbf{t}|\boldsymbol{\Phi}\boldsymbol{w}, \boldsymbol{\Omega})$ with covariance of measurements as $\Omega_{ij} = Cov(t_i, t_j)$
- prior: $p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$

Task at hand:
Derive the MAP solution by reducing it to the uncorrelated case via a change to the eigenbasis of the covariance matrix.

a) Write down the likelihood $p(\mathcal{D}|\boldsymbol{\theta})$ in vector/matrix form, i.e. in terms of $\mathbf{t}, \boldsymbol{\Phi}, \boldsymbol{w}$ and $\boldsymbol{\Omega}$.

As the samples are no longer independent, the product of univariate Gaussians does not longer apply. Now, a multivariate Gaussian with respective density function is needed. Inserting the appropriate μ and covariance terms into the density function yields:

$$p(\mathcal{D}|\boldsymbol{\theta}) = p(\mathbf{t}|\boldsymbol{\Phi}, \boldsymbol{w}, \boldsymbol{\Omega}) = \mathcal{N}(\mathbf{t}|\boldsymbol{\Phi}\boldsymbol{w}, \boldsymbol{\Omega}) = \frac{1}{(2\pi)^{\frac{N}{2}}|\boldsymbol{\Omega}|^{\frac{1}{2}}} \cdot exp\left(-\frac{1}{2}(\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{w})^T \boldsymbol{\Omega}^{-1}(\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{w})\right)$$

b) Write the likelihood in terms of a Gaussian distribution with a diagonal covariance matrix by changing the basis of the space in which the targets are expressed. Specifically express the covariance matrix in its eigenbasis, i.e. write it as $\boldsymbol{\Omega} = \mathbf{A}^T\mathbf{D}\mathbf{A}$, with $\mathbf{D}$ being a diagonal matrix containing the eigenvalues of $\boldsymbol{\Omega}$ and $\mathbf{A}^T = \mathbf{A}^{-1}$ being an orthogonal change of basis.

At first the term in the exponent is taken care of and expressed in the sought form. For that,

$$\begin{aligned}
\boldsymbol{\Omega} &= \mathbf{A}^T\mathbf{D}\mathbf{A} \\
\Rightarrow \boldsymbol{\Omega}^{-1} &= (\mathbf{A}^T\mathbf{D}\mathbf{A})^{-1} \\
&= \mathbf{A}^{-1}\mathbf{D}^{-1}\mathbf{A}^{T^{-1}} \\
&= \mathbf{A}^T\mathbf{D}^{-1}\mathbf{A}
\end{aligned}$$

and as hinted in the sheet $\boldsymbol{\tau} := \mathbf{A}\mathbf{t}$ and $\boldsymbol{\Psi} := \mathbf{A}\boldsymbol{\Phi}$ are used. So:

$$\begin{aligned}
(\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{w})^T\boldsymbol{\Omega}^{-1}(\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{w}) &= \mathbf{t}^T\boldsymbol{\Omega}^{-1}\mathbf{t} - \mathbf{t}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}\boldsymbol{w} - \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\Phi}^{\boldsymbol{T}}\boldsymbol{\Omega}^{\boldsymbol{-1}}\mathbf{t} + \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\Phi}^{\boldsymbol{T}}\boldsymbol{\Omega}^{\boldsymbol{-1}}\boldsymbol{\Phi}\boldsymbol{w} \\
&= \mathbf{t}^T\boldsymbol{A}^T\mathbf{D}^{-1}\boldsymbol{A}\mathbf{t} - 2\mathbf{t}^T\boldsymbol{A}^T\mathbf{D}^{-1}\boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{w}^T\boldsymbol{\Phi}^T\boldsymbol{A}^T\mathbf{D}^{-1}\boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{w} \\
&= \boldsymbol{\tau}^T\mathbf{D}^{-1}\boldsymbol{\tau} - 2\boldsymbol{\tau}^T\mathbf{D}^{-1}\boldsymbol{\Psi}\boldsymbol{w} + (\boldsymbol{\Psi}\boldsymbol{w})^T\mathbf{D}^{-1}\boldsymbol{\Psi}\boldsymbol{w} \\
&= (\boldsymbol{\tau} - \boldsymbol{\Psi}\boldsymbol{w})^T\mathbf{D}^{-1}(\boldsymbol{\tau} - \boldsymbol{\Psi}\boldsymbol{w})
\end{aligned}$$

For the normalizing factor, we can use the orthogonality of matrix **A**(orthogonal matrix **A**: $\mathbf{A}\mathbf{A}^T = \mathbf{I}$ and $det(\mathbf{I}) = 1$):

$$
\begin{aligned}
|\boldsymbol{\Omega}| = |\boldsymbol{A}^T\mathbf{D}\boldsymbol{A}| = det(\boldsymbol{A}^T\mathbf{D}\boldsymbol{A}) = det(\mathbf{A}^T)det(\mathbf{D})det(\mathbf{A}) = det(\mathbf{A}^T)det(\mathbf{A})det(\mathbf{D}),\\
= det(\mathbf{A}\mathbf{A}^T)det(\mathbf{D}) = det(\mathbf{I})det(\mathbf{D}) = det(\mathbf{D}) = |\mathbf{D}|
\end{aligned}
$$

, whereas λ_i denote the eigenvalues of matrix D and as it is a diagonal matrix, its determinant is the product of the eigenvalues. Assembling those properties, the according likelihood arises as:

$$
\begin{aligned}
p(\mathcal{D}|\boldsymbol{\theta}) &= \frac{1}{(2\pi)^{\frac{N}{2}}|\boldsymbol{D}|^{\frac{1}{2}}} \cdot exp\left(-\frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{\Psi}\boldsymbol{w})^T\boldsymbol{D}^{-1}(\boldsymbol{\tau} - \boldsymbol{\Psi}\boldsymbol{w})\right)\\
&= \mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Psi}\boldsymbol{w}, \boldsymbol{D})
\end{aligned}
$$

c) Factorize the distribution into a product of univariate Gaussians. Write ψ_i for the i-th row of the matrix Ψ. λ_i shall denote the i-th entry of the Diagonal matrix D, representing the i-th eigenvalue of $\boldsymbol{\Omega}$.

$$
\begin{aligned}
\mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Psi}\boldsymbol{w}, \boldsymbol{D}) &= \frac{1}{(2\pi)^{\frac{N}{2}}|\boldsymbol{D}|^{\frac{1}{2}}} \cdot exp\left(-\frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{\Psi}\boldsymbol{w})^T\boldsymbol{D}^{-1}(\boldsymbol{\tau} - \boldsymbol{\Psi}\boldsymbol{w})\right)\\
&= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\lambda_i}} \cdot exp\left(-\frac{1}{2\lambda_i}(\tau_i - \psi_i\boldsymbol{w})^2\right)\\
&= \prod_{i=1}^{N} \mathcal{N}(\tau_i|\psi_i\boldsymbol{w}, \lambda_i)
\end{aligned}
$$

d) Write down the explicit form of the prior $p(\boldsymbol{w})$, i.e. use the expression for a multivariate Gaussian distribution with the correct mean and covariance. Compute the logarithm of the prior $log(p(\boldsymbol{w})$.

Taking the prior as specified, $p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$ holds. That way the explicit form emerges as:

$$
\begin{aligned}
p(\boldsymbol{w}) &= \frac{1}{(2\pi)^{\frac{M}{2}}|\alpha^{-1}\mathbf{I}|^{\frac{1}{2}}} \cdot exp\left(-\frac{1}{2}(\boldsymbol{w}^T(\alpha^{-1}\mathbf{I})^{-1}\boldsymbol{w})\right)\\
&= \frac{\alpha^{\frac{M}{2}}}{(2\pi)^{\frac{M}{2}}} \cdot exp\left(-\frac{\alpha}{2}(\boldsymbol{w}^T\boldsymbol{w})\right)\\
&= \prod_{i=0}^{M-1}\left(\frac{\alpha}{(2\pi)}\right)^{\frac{1}{2}} \cdot exp\left(-\frac{\alpha}{2}w_i^2\right),
\end{aligned}
$$

with $|\alpha^{-1}\mathbf{I}|^{\frac{1}{2}} = \sqrt{\left(\frac{1}{\alpha}\right)^M} = \left(\frac{1}{\alpha}\right)^{\frac{M}{2}}$ due to the diagonality of the identity matrix. Therefore, the log prior is as follows:

$$
log\ p(\boldsymbol{w}) = \frac{M}{2}log(\alpha) - \frac{M}{2}log(2\pi) - \frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w}
$$

e) Write down an expresion for the posterior $p(\boldsymbol{w}|\mathcal{D}) = p(\boldsymbol{w}|\boldsymbol{\Psi}, \boldsymbol{\tau}, \mathbf{D})$ over $\boldsymbol{w}$ by applying Bayes rule. The explicit form is not needed.

$$
\begin{aligned}
p(\boldsymbol{w}|\mathcal{D}) &= \frac{p(\mathcal{D}|\boldsymbol{w}) \cdot p(\boldsymbol{w})}{p(\mathcal{D})} = \frac{p(\mathcal{D}|\boldsymbol{w}) \cdot p(\boldsymbol{w})}{\int p(\mathcal{D}|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}}\\
&= \frac{p(\boldsymbol{\tau}|\boldsymbol{w}, \boldsymbol{\Psi}, \boldsymbol{D}) \cdot p(\boldsymbol{w})}{\int p(\boldsymbol{\tau}|\boldsymbol{w}, \boldsymbol{\Psi}, \boldsymbol{D})p(\boldsymbol{w})d\boldsymbol{w}}\\
&= \frac{\mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Psi}\boldsymbol{w}, \mathbf{D}) \cdot \mathcal{N}(\boldsymbol{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})}{\int \mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Psi}\boldsymbol{w}, \mathbf{D}) \cdot \mathcal{N}(\boldsymbol{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})d\boldsymbol{w}}
\end{aligned}
$$

The denominator consists of an integral, which is hardly computable and thus requires means of approximation.

f) Compute the log-posterior for both the matrix form of the likelihood as derived in 1.2 and the factorized component form of the likelihood as derived in 1.3. Collect all terms which are independent of **w** into a constant **C**.

$$log(p(\boldsymbol{w}|\boldsymbol{\Psi},\boldsymbol{\tau},\boldsymbol{D})) = log(p(\boldsymbol{\tau}|\boldsymbol{w},\boldsymbol{\Psi},\boldsymbol{D})) + log(p(\boldsymbol{w})) - log(p(\mathcal{D}))$$

At first regarding 1.2 in the matrix form:

$$\begin{aligned} log(p(\boldsymbol{w}|\boldsymbol{\Psi},\boldsymbol{\tau},\boldsymbol{D}) &= log\left(\frac{1}{(2\pi)^{\frac{N}{2}}|\boldsymbol{D}|^{\frac{1}{2}}}\cdot exp\left(-\frac{1}{2}(\boldsymbol{\tau}-\boldsymbol{\Psi}\boldsymbol{w})^T\boldsymbol{D}^{-1}(\boldsymbol{\tau}-\boldsymbol{\Psi}\boldsymbol{w})\right)\right) \\ &+ \frac{M}{2}log(\alpha) - \frac{M}{2}ln(2\pi) - \frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w} \\ &- log(p(\mathcal{D})) \\ &= -\frac{1}{2}\left((\boldsymbol{\tau}-\boldsymbol{\Psi}\boldsymbol{w})^T\boldsymbol{D}^{-1}(\boldsymbol{\tau}-\boldsymbol{\Psi}\boldsymbol{w})\right) - \frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w} + c, \end{aligned}$$

whereas

$$c = -\frac{N}{2}log(2\pi) - \frac{1}{2}log(|\mathbf{D}|) + \frac{M}{2}log(\alpha) - \frac{M}{2}ln(2\pi) - log(p(\mathcal{D}))$$

Same holds for the explicit form:

$$\begin{aligned} log(p(\boldsymbol{w}|\boldsymbol{\Psi},\boldsymbol{\tau},\boldsymbol{D}) &= log\left(\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\lambda_i}}\cdot exp\left(-\frac{1}{2\lambda_i}(\tau_i-\psi_i\boldsymbol{w})^2\right)\right) + log\left(\prod_{i=0}^{M-1}\left(\frac{\alpha}{(2\pi)}\right)^{\frac{1}{2}}\cdot exp\left(-\frac{\alpha}{2}w_i^2\right)\right) - log(p(\mathcal{D})) \\ &= -\sum_{i=1}^{N}\frac{1}{2\lambda_i}(\tau_i-\psi_i\boldsymbol{w})^2 - \sum_{j=0}^{M-1}\frac{\alpha}{2}w_j^2 + c, \end{aligned}$$

whereas again c consists of the terms

$$c = -\frac{N}{2}log(2\pi) - \frac{1}{2}log\left(\sum_{i=1}^{N}log(\lambda_i)\right) + \frac{M}{2}log(\alpha) - \frac{M}{2}log(2\pi) - log(p(\mathcal{D}))$$

g) Solve for $\mathbf{w}_{MAP}$ by first taking the derivative of the log-posterior with respect to **w**, then settling it to 0 and finally solving for **w**.

We will draw on the matrix form just derived. Therefore,

$$\begin{aligned} log(p(\boldsymbol{w}|\boldsymbol{\Psi},\boldsymbol{\tau},\boldsymbol{D}) &= -\frac{1}{2}\left(\boldsymbol{\tau}^T\mathbf{D}^{-1}\boldsymbol{\tau} - 2\boldsymbol{\tau}^T\mathbf{D}^{-1}\boldsymbol{\Psi}\boldsymbol{w} + (\boldsymbol{\Psi}\boldsymbol{w})^T\mathbf{D}^{-1}\boldsymbol{\Psi}\boldsymbol{w}\right) - \frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w} + c \\ \frac{\partial}{\partial\boldsymbol{w}}log(p(\boldsymbol{w}|\boldsymbol{\Psi},\boldsymbol{\tau},\boldsymbol{D}) &= \boldsymbol{\tau}^T\mathbf{D}^{-1}\boldsymbol{\Psi} - \boldsymbol{w}^T\boldsymbol{\Psi}^T\mathbf{D}^{-1}\boldsymbol{\Psi} - \alpha\boldsymbol{w}^T \overset{!}{=} 0 \\ \Rightarrow \quad \boldsymbol{\tau}^T\mathbf{D}^{-1}\boldsymbol{\Psi} &= \alpha\boldsymbol{w}^T + \boldsymbol{w}^T\boldsymbol{\Psi}^T\mathbf{D}^{-1}\boldsymbol{\Psi} \\ \Rightarrow \quad \boldsymbol{\tau}^T\mathbf{D}^{-1}\boldsymbol{\Psi} &= \boldsymbol{w}^T(\alpha\mathbf{I} + \boldsymbol{\Psi}^T\mathbf{D}^{-1}\boldsymbol{\Psi}) \\ \text{transpose on both sides} \Rightarrow \quad \boldsymbol{\Psi}^T\mathbf{D}^{-1}\boldsymbol{\tau} &= (\alpha\mathbf{I} + \boldsymbol{\Psi}^T\mathbf{D}^{-1}\boldsymbol{\Psi})\boldsymbol{w} \\ \Rightarrow \quad \boldsymbol{w}_{MAP} &= (\alpha\mathbf{I} + \boldsymbol{\Psi}^T\mathbf{D}^{-1}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^T\mathbf{D}^{-1}\boldsymbol{\tau} \end{aligned}$$

h) Express the solutions $\mathbf{w}_{MAP}$ in terms of the original quantities t and $\boldsymbol{\Psi}$ to end up with the final solution stated above.

Using, $\boldsymbol{\Psi} = \mathbf{A\Phi}$ and $\boldsymbol{\tau} = \mathbf{At}$ we arrive at the MAP estimate for $\boldsymbol{w}$ in terms of the original quantities:

$$\begin{aligned}\boldsymbol{w}_{MAP} &= (\alpha\mathbf{I} + (\mathbf{A\Phi})^T\mathbf{D}^{-1}\mathbf{A\Phi})^{-1}(\mathbf{A\Phi})^T\mathbf{D}^{-1}\mathbf{At}\\ &= (\alpha\mathbf{I} + \mathbf{\Phi}^T\mathbf{\Omega}^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{\Omega}^{-1}\mathbf{t},\end{aligned}$$

which is exactly the form given as final solution on the homework sheet.

2 ML estimate of angle measurements

Find the ML estimate of the angle θ, given that you have two independent noisy measurements, c and s, where c is a measire pf the cosine and s is a measure of the sine, of the angle θ. Assume each measurement has a known Gaussian standard deviation δ
As c and s are measurements with a Gaussian noise around cosin and respectively sin of θ, we know the following:

- $c \sim \mathcal{N}(cos(\theta), \delta^2)$, with according density: $p(c|\theta) = \frac{1}{(2\pi\delta^2)^{\frac{1}{2}}} \cdot exp\left(-\frac{1}{2\delta^2}(c - cos(\theta))^2\right)$
- $s \sim \mathcal{N}(sin(\theta), \delta^2)$, with according density: $p(s|\theta) = \frac{1}{(2\pi\delta^2)^{\frac{1}{2}}} \cdot exp\left(-\frac{1}{2\delta^2}(s - sin(\theta))^2\right)$

The likelihood of the joint instance of both c and s is with respect two their independence:

$$\begin{aligned}p(c, s|\theta) &= p(c|\theta) \cdot p(s|\theta)\\ log(p(c, s|\theta)) &= log(p(c|\theta)) + log(p(s|\theta))\end{aligned}$$

As always, the log-likelihood is derived and set zero:

$$\begin{aligned}\frac{\partial}{\partial\theta}log(p(c, s|\theta) &= \frac{\partial}{\partial\theta}log\left(\frac{1}{(2\pi\delta^2)^{\frac{1}{2}}} \cdot exp\left(-\frac{1}{2\delta^2}(c - cos(\theta))^2\right) \cdot \frac{1}{(2\pi\delta^2)^{\frac{1}{2}}} \cdot exp\left(-\frac{1}{2\delta^2}(s - sin(\theta))^2\right)\right)\\ &= \frac{\partial}{\partial\theta}log\left(\underbrace{\left(\frac{1}{2\pi\delta^2}\right)}_{\text{independent of }\theta} - \left(\frac{1}{2\delta^2}((c - cos(\theta))^2 + (s - sin(\theta))^2\right)\right)\\ \text{(using } sin(\theta)^2 + cos(\theta)^2 = 1\text{)} \quad &= \frac{\partial}{\partial\theta} - \left(\frac{c^2 - 2c \cdot cos(\theta) + cos(\theta)^2 + s^2 - 2s \cdot sin(\theta) + sin(\theta)^2}{2\delta^2}\right)\\ &= -\frac{\partial}{\partial\theta}\left(\underbrace{\frac{c^2 + s^2 + 1}{2\delta^2}}_{:=C\text{ independent of }\theta} + \frac{-2c \cdot cos(\theta) - 2s \cdot sin(\theta)}{2\delta^2}\right)\\ &= \frac{1}{2\delta^2}\frac{\partial}{\partial\theta}2c \cdot cos(\theta) + 2s \cdot sin(\theta)\\ &= \frac{1}{2\delta^2}(2c \cdot -sin(\theta) + 2s \cdot cos(\theta)) \overset{!}{=} 0\end{aligned}$$

$$\begin{aligned}&\Rightarrow \quad s \cdot cos(\theta) = c \cdot sin(\theta)\\ &\Rightarrow \quad \frac{s}{c} = \frac{sin(\theta)}{cos(\theta)}\\ &\Rightarrow \quad \frac{s}{c} = tan(\theta)\\ &\Rightarrow \quad tan\left(\frac{s}{c}\right)^{-1} = \theta_{ML}\end{aligned}$$

3 ML and MAP solution of a Poisson distribution fit

Suppose we have a set of N integer non-negative measurements $x_1, \ldots, x_N$ from a stochastic process.

a) Demonstrate analytically that the ml fit of a Poisson distribution to this data is givne by $\lambda_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i$.

We assume iid samples x_i with $x_i \sim Poisson(\lambda)$. Therefore the likelihood of the sample set is given by:

$$\begin{aligned} p(x_1, \dots, x_N|\lambda) &= \prod_{i=1}^{N} p(x_i|\lambda) \\ &= \prod_{i=1}^{N} \frac{\lambda^{x_i}}{x_i!} \cdot exp(-\lambda) \\ &= \frac{\lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i!} exp(-N\lambda) \end{aligned}$$

Again, setting the log-likelihood equal to zero yields:

$$\begin{aligned} \frac{\partial}{\partial\lambda} log(p(x_1, \dots, x_N|\lambda)) &= \frac{\partial}{\partial\lambda} \sum_{i=1}^{N} x_i \cdot log(\lambda) - \sum_{i=1}^{N} log(x_i!) - N\lambda \\ &= \frac{1}{\lambda} \cdot \sum_{i=1}^{N} x_i - N \overset{!}{=} 0 \\ \Rightarrow \quad \lambda_{ML} &= \frac{1}{N} \sum_{i=1}^{N} x_i \end{aligned}$$

b) We are then given additional prior information that $p(\lambda) \propto exp(-\lambda/a)$ for $\lambda \geq 0$ and zero otherwise, for a known value $a > 0$. What is the maximum posterior solution for λ?

Let $x_1, \dots, x_N = \mathbf{x}$, then the posterior of λ is:

$$\begin{aligned} p(\lambda|\mathbf{x}) &= \frac{p(\mathbf{x}|\lambda) \cdot p(\lambda)}{p(\mathbf{x})} \\ \text{log-likelihood:} \quad log(p(\lambda|\mathbf{x})) &= log(p(\mathbf{x}|\lambda)) + log(p(\lambda)) - log(p(\mathbf{x})) \\ log(p(\lambda|\mathbf{x})) &= \frac{\partial}{\partial\lambda} \sum_{i=1}^{N} x_i \cdot log(\lambda) - \sum_{i=1}^{N} log(x_i!) - N\lambda - \frac{\lambda}{a} - log(p(\mathbf{x})) \\ \Rightarrow \quad \frac{\partial}{\partial\lambda} &= \frac{1}{\lambda} \cdot \sum_{i=1}^{N} x_i - N - \frac{1}{a} \overset{!}{=} 0 \\ \Rightarrow \quad \lambda_{MAP} &= \frac{1}{N + \frac{1}{a}} \sum_{i=1}^{N} x_i \end{aligned}$$

c) What is the effect of the prior on the value of the estimate, (compared to the maximum likelihood case) and what happens in the limits $a \to \infty$ and $a \to 0$?

The effect is that in the denominator before the sum over all x_i, there is not only N but also $\frac{1}{a}$ added.

For $a \to \infty$, $\frac{1}{a} \to 0$, that is why, $\lambda_{MAP} \to \frac{1}{N}\sum_{i=1}^{N} x_i$. So for $a \to \infty$, the MAP goes towards the ML estimate.

For $a \to 0$, $\frac{1}{a} \to \infty$, that is why, $\lambda_{MAP} \to \frac{1}{'\infty'}\sum_{i=1}^{N} x_i \to 0$. So for $a \to 0$, the MAP goes towards 0.