Exercise

03

TUM Department of Informatics

Supervised by Prof. Dr. Stephan Günnemann
Informatics 3 - Professorship of Data Mining and Analytics

Submitted by Marcel Bruckner (03674122)
Julian Hohenadel (03673879)
Kevin Bein (03707775)

Submission date Munich, October 26, 2019

Optimizing Likelihoods: Monotonic Transforms

Problem 1:

$$f_1(\theta) = \theta^t(1-\theta)^h$$

$$f_2(\theta) = \log(\theta^t(1-\theta)^h)$$

$$\begin{aligned}
\frac{\partial}{\partial\theta} f_1(\theta) &= t\theta^{t-1}(1-\theta)^h + \theta^t \frac{\partial}{\partial\theta}\left((1-\theta)^h\right) \\
&= t\theta^{t-1}(1-\theta)^h - \theta^t h(1-\theta)^{h-1} \\
\frac{\partial^2}{\partial\theta} f_1(\theta) &= t \cdot \frac{\partial}{\partial\theta}\left(\theta^{t-1}(1-\theta)^h\right) - h \cdot \frac{\partial}{\partial\theta}\left(\theta^t(1-\theta)^{h-1}\right) \\
&= t \cdot \left((t-1)\theta^{t-2}(1-\theta)^h - \theta^{t-1}h(1-\theta)^{h-1}\right) \\
&+ h \cdot \left(t\theta^{t-1}(1-\theta)^{h-1} - \theta^t(h-1)(1-\theta)^{h-2}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\theta} f_2(\theta) &= \frac{\partial}{\partial\theta}\left(\log(\theta^t) + \log((1-\theta)^h)\right) \\
&= \frac{\partial}{\partial\theta}\left(\log(\theta^t)\right) + \frac{\partial}{\partial\theta}\left(\log((1-\theta)^h)\right) \\
&= t \cdot \frac{\partial}{\partial\theta}(\theta) + h \cdot \frac{\partial}{\partial\theta}(\log(1-\theta)) \\
&= \frac{t}{\theta} - \frac{h}{1-\theta}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial\theta} f_2(\theta) &= \frac{\partial}{\partial\theta}\left(\frac{t}{\theta}\right) - \frac{\partial}{\partial\theta}\left(\frac{h}{1-\theta}\right) \\
&= \frac{0\cdot\theta - t\cdot 1}{\theta^2} - \frac{0\cdot(1-\theta) - h\cdot(-1)}{(1-\theta)^2} \\
&= -\frac{t}{\theta^2} - \frac{h}{(1-\theta)^2}
\end{aligned}$$

Problem 2:

"Monotonic functions preserver critical points"

Since $\log$ is a monotonic transformation, the function will yield the same values when plugged into $\arg\max_\theta$:

$$\arg\max_\theta f(\theta) = \arg\max_\theta \log f(\theta)$$

Considering problem 1 and the fact that log converts exponents to factors (which also increases the numerical stabillity), computing the log-likelihood is faster and yields more compact functions which can then be maximized.

Properties of MLE and MAP

Problem 3:

$$\begin{aligned}
Beta(6,4) &= \left(\frac{\Gamma(6+4)}{\Gamma(6)\Gamma(4)} \cdot \theta^{6-1} \cdot (1-\theta)^{4-1}\right) \\
&= \frac{9!}{5! \cdot 4!} \cdot \theta^5 \cdot (1-\theta)^3 \\
&= 126 \cdot \theta^5 \cdot (1-\theta)^3
\end{aligned}$$

$$\begin{aligned}
\theta_{\text{MAP}} &= \arg\max_\theta p(\theta|f) \\
&= \arg\max_\theta \frac{p(f|\theta) \cdot p(\theta)}{p(f)} \\
&= \arg\max_\theta p(f|\theta) \cdot p(\theta) \\
&= \arg\max_\theta \left(\theta^{\mathbb{I}[f=T]} \cdot (1-\theta)^{\mathbb{I}[f=H]}\right) \cdot \left(126 \cdot \theta^5 \cdot (1-\theta)^3\right) \\
&= \arg\max_\theta \theta^{M+5} \cdot (1-\theta)^{N+3} \\
&= \arg\max_\theta \log\left(\theta^{M+5} \cdot (1-\theta)^{N+3}\right) \\
&= \arg\max_\theta (M+5)\log(\theta) + (N+3)\log(1-\theta)
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \frac{\partial}{\partial\theta}(M+5)\log(\theta) + (N+3)\log(1-\theta) = \frac{M+5}{\theta} - \frac{N-3}{1-\theta} \stackrel{!}{=} 0 \\
&\Rightarrow \theta_{\text{MAP}} = \frac{M+5}{M+N+8}
\end{aligned}$$

$$\begin{aligned}
\theta_{\text{MAP}} \stackrel{!}{=} 0.75 = \frac{3}{4} = \frac{30}{40} \\
\Rightarrow \quad M = 25, N = 40 - 25 - 8 = 7
\end{aligned}$$

(In general, the solution is $M = 3N + 1$)

Problem 4:

Programming Task

Problem 5:

Appendix

We confirm that the submitted solution is original work and was written by us without further assistance. Appropriate credit has been given where reference has been made to the work of others.

Munich, October 26, 2019, Signature Marcel Bruckner (03674122)

Munich, October 26, 2019, Signature Julian Hohenadel (03673879)

Munich, October 26, 2019, Signature Kevin Bein (03707775)