

# Notes on Pattern Recognition and Machine Learning

Max Leung

April 11, 2023

## 1 Laplace Approximation

$$\begin{aligned} \ln f(\mathbf{z}) &\approx \ln f(\mathbf{z}_0) + \overbrace{\nabla \ln f(\mathbf{z}_0)'}^{\mathbf{o}'} (\mathbf{z} - \mathbf{z}_0) + \frac{1}{2} (\mathbf{z} - \mathbf{z}_0)' \overbrace{\nabla \nabla \ln f(\mathbf{z}_0)}^{-\mathbf{A}} (\mathbf{z} - \mathbf{z}_0) \\ &= \ln f(\mathbf{z}_0) - \frac{1}{2} (\mathbf{z} - \mathbf{z}_0)' \mathbf{A} (\mathbf{z} - \mathbf{z}_0) \end{aligned}$$

$$\begin{aligned} f(\mathbf{z}) &\approx \exp(\ln f(\mathbf{z}_0) - \frac{1}{2} (\mathbf{z} - \mathbf{z}_0)' \mathbf{A} (\mathbf{z} - \mathbf{z}_0)) \\ &= \exp(\ln f(\mathbf{z}_0)) \exp(-\frac{1}{2} (\mathbf{z} - \mathbf{z}_0)' \mathbf{A} (\mathbf{z} - \mathbf{z}_0)) \\ &= f(\mathbf{z}_0) \exp(-\frac{1}{2} (\mathbf{z} - \mathbf{z}_0)' \mathbf{A} (\mathbf{z} - \mathbf{z}_0)) \end{aligned}$$

If we approximate  $f(\cdot)$  by  $N(\mathbf{z}_0, \mathbf{A}^{-1})$ , we have

$$\begin{aligned} &\approx \frac{1}{(2\pi)^{M/2} |\mathbf{A}|^{-1/2}} \underbrace{\exp\{-\frac{1}{2} (\mathbf{z}_0 - \mathbf{z}_0)' \mathbf{A} (\mathbf{z}_0 - \mathbf{z}_0)\}}_1 \exp(-\frac{1}{2} (\mathbf{z} - \mathbf{z}_0)' \mathbf{A} (\mathbf{z} - \mathbf{z}_0)) \\ &= N(\mathbf{z} | \mathbf{z}_0, \mathbf{A}^{-1}) = q(\mathbf{z}) \end{aligned}$$

where  $\mathbf{z}_0 = \arg \max_{\mathbf{z}} \ln f(\mathbf{z})$  and  $\mathbf{A} = -\nabla \nabla \ln f(\mathbf{z}_0)$

## 2 Bayesian Logistic Regression

### 2.1 Posterior Distribution of Parameters

Assume we have prior density  $p(\mathbf{w}) = N(\mathbf{w} | \mathbf{m}_0, \mathbf{S}_0)$

Likelihood function is  $p(\mathbf{t} | \mathbf{w}) = \prod_{n=1}^N p(C_1 | \mathbf{x}_n; \mathbf{w})^{t_n} (1 - p(C_1 | \mathbf{x}_n; \mathbf{w}))^{1-t_n}$  where  $p(C_1 | \mathbf{x}_n; \mathbf{w}) = \sigma(\mathbf{w}' \mathbf{x}_n)$

Posterior density is

$$\begin{aligned} p(\mathbf{w} | \mathbf{t}) &\propto p(\mathbf{t} | \mathbf{w}) p(\mathbf{w}) \\ &= \prod_{n=1}^N \sigma(\mathbf{w}' \mathbf{x}_n)^{t_n} (1 - \sigma(\mathbf{w}' \mathbf{x}_n))^{1-t_n} N(\mathbf{w} | \mathbf{m}_0, \mathbf{S}_0) \end{aligned}$$

which is not a well known joint density function

$$\begin{aligned} \ln p(\mathbf{w} | \mathbf{t}) &= \ln \left[ \prod_{n=1}^N \sigma(\mathbf{w}' \mathbf{x}_n)^{t_n} (1 - \sigma(\mathbf{w}' \mathbf{x}_n))^{1-t_n} N(\mathbf{w} | \mathbf{m}_0, \mathbf{S}_0) \right] \\ &= \sum_{n=1}^N [t_n \ln \sigma(\mathbf{w}' \mathbf{x}_n) + (1 - t_n) \ln (1 - \sigma(\mathbf{w}' \mathbf{x}_n))] + \ln [N(\mathbf{w} | \mathbf{m}_0, \mathbf{S}_0)] \\ &= \sum_{n=1}^N [t_n \ln \sigma(\mathbf{w}' \mathbf{x}_n) + (1 - t_n) \ln (1 - \sigma(\mathbf{w}' \mathbf{x}_n))] + \ln \left[ \frac{1}{(2\pi)^{D/2} |\mathbf{S}_0|^{1/2}} \exp\left\{-\frac{1}{2} (\mathbf{w} - \mathbf{m}_0)' \mathbf{S}_0^{-1} (\mathbf{w} - \mathbf{m}_0)\right\} \right] \\ &= \sum_{n=1}^N [t_n \ln \sigma(\mathbf{w}' \mathbf{x}_n) + (1 - t_n) \ln (1 - \sigma(\mathbf{w}' \mathbf{x}_n))] + \ln \left[ \frac{1}{(2\pi)^{D/2} |\mathbf{S}_0|^{1/2}} \right] - \frac{1}{2} (\mathbf{w} - \mathbf{m}_0)' \mathbf{S}_0^{-1} (\mathbf{w} - \mathbf{m}_0) \end{aligned}$$

We can approximate  $p(\mathbf{w}|\mathbf{t})$  by Laplace Approximation. As a result, our posterior follows multivariate normal distribution.

$$p(\mathbf{w}|\mathbf{t}) \approx q(\mathbf{w}) = N(\mathbf{w}|\mathbf{w}_{MAP}, \mathbf{S}^{-1})$$

where  $\mathbf{w}_{MAP} = \arg \max_{\mathbf{w}} \ln p(\mathbf{w}|\mathbf{t})$

$$\begin{aligned} \frac{\partial \ln p(\mathbf{w}|\mathbf{t})}{\partial \mathbf{w}} \Big|_{\mathbf{w}_{MAP}} &= \mathbf{0} \\ \frac{\partial \sum_{n=1}^N [t_n \ln \sigma(\mathbf{w}'\mathbf{x}) + (1 - t_n) \ln(1 - \sigma(\mathbf{w}'\mathbf{x}))] + \ln[\frac{1}{(2\pi)^{D/2} |\mathbf{S}_0|^{1/2}}] - \frac{1}{2}(\mathbf{w} - \mathbf{m}_0)' \mathbf{S}_0^{-1}(\mathbf{w} - \mathbf{m}_0)}{\partial \mathbf{w}} \Big|_{\mathbf{w}_{MAP}} &= \mathbf{0} \\ \mathbf{X}'(\mathbf{t} - \mathbf{p}) - \mathbf{S}_0^{-1}(\mathbf{w}_{MAP} - \mathbf{m}_0) &= \mathbf{0} \end{aligned}$$

where  $\mathbf{p} = (\sigma(\mathbf{w}'_{MAP}\mathbf{x}_1), \dots, \sigma(\mathbf{w}'_{MAP}\mathbf{x}_D))'$

There is no closed form solution for  $\mathbf{w}_{MAP}$

$$\begin{aligned} \mathbf{S} &= -\nabla \nabla \ln p(\mathbf{w}_{MAP}|\mathbf{t}) \\ &= -\nabla \nabla \left\{ \sum_{n=1}^N [t_n \ln \sigma(\mathbf{w}'\mathbf{x}) + (1 - t_n) \ln(1 - \sigma(\mathbf{w}'\mathbf{x}))] + \ln[\frac{1}{(2\pi)^{D/2} |\mathbf{S}_0|^{1/2}}] - \frac{1}{2}(\mathbf{w} - \mathbf{m}_0)' \mathbf{S}_0^{-1}(\mathbf{w} - \mathbf{m}_0) \right\} \\ &= -(-\mathbf{X}'\mathbf{W}\mathbf{X} - \mathbf{S}_0^{-1}) \\ &= \mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{S}_0^{-1} \end{aligned}$$

where  $(\mathbf{W})_{ii} = \sigma(\mathbf{w}'_{MAP}\mathbf{x}_i)(1 - \sigma(\mathbf{w}'_{MAP}\mathbf{x}_i))$  and  $(\mathbf{W})_{ij} = 0$  for  $i \neq j$

### 3 Reference

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. New York :Springer.