

Bayesian logistic regression Lecture 10 K=2 here

Assume $p(\vec{w}) = \mathcal{N}(\vec{w} | \vec{0}, \underbrace{\alpha^{-1} \mathbb{I}}_{\text{"S}_0"})$

$$p(\vec{w} | \vec{E}) \sim p(\vec{E} | \vec{w}) p(\vec{w})$$

Then $\log p(\vec{w} | \vec{E}) = -\frac{\alpha}{2} \vec{w}^T \vec{w} + \sum_{n=1}^N [t_n \log y_n + (1-t_n) \log(1-y_n)] + \text{const}(\vec{w}),$
(*)

where $y_n = \sigma(\vec{w}^T \cdot \vec{y}_n)$ $t_n = \{0, 1\}$

(*) is at max when $\vec{w} = \vec{w}_{\text{MAP}}$; then

$$S_N \equiv -\nabla_{\vec{w}} \nabla_{\vec{w}} \log p(\vec{w} | \vec{E}) = \alpha \mathbb{I} + \sum_{n=1}^N y_n(1-y_n) \vec{y}_n \vec{y}_n^T$$

Thus $p(\vec{w} | \vec{E}) \Rightarrow q(\vec{w} | \vec{E}) = \mathcal{N}(\vec{w} | \vec{w}_{\text{MAP}}, S_N)$.

Next, predictive distribution:

$$\left\{ \begin{aligned} p(c_1 | \vec{y}, \vec{E}) &= \int d\vec{w} p(c_1 | \vec{y}, \vec{w}) p(\vec{w} | \vec{E}) = \\ &= \int d\vec{w} \underbrace{\sigma(\vec{w}^T \vec{y})}_{\text{"a"}} q(\vec{w} | \vec{E}) \\ p(c_2 | \vec{y}, \vec{E}) &= 1 - p(c_1 | \vec{y}, \vec{E}) \end{aligned} \right.$$

here input vector of features

Use $\sigma(\vec{w}^T \vec{y}) = \int da \sigma(a) \delta(a - \vec{w}^T \vec{y})$ to obtain:

since $p(\vec{w} | \vec{E})$ replaced by $q(\vec{w} | \vec{E})$

$$p(c_1 | \vec{y}, \vec{E}) \approx \int da \sigma(a) p(a), \text{ where}$$

$$p(a) = \int d\vec{w} q(\vec{w} | \vec{E}) \delta(a - \vec{w}^T \vec{y})$$

So, compute

$$P(a) = \frac{1}{(2\pi)^{M/2}} \frac{1}{|S_N|^{1/2}} \int d\vec{w} \frac{1}{2\pi} \int dk e^{ik(a - \vec{w}^T \vec{g})_x} \\ \times e^{-\frac{1}{2} (\vec{w} - \vec{w}_{MAP})^T S_N^{-1} (\vec{w} - \vec{w}_{MAP})} \quad \text{⊖}$$

↑
 $\vec{w}' = \vec{w} - \vec{w}_{MAP}$

$$\text{⊖} \quad \frac{1}{(2\pi)^{M/2}} \frac{1}{|S_N|^{1/2}} \frac{1}{2\pi} \int dk e^{ik(a - \vec{g}^T \vec{w}_{MAP})_x} \\ \times \int d\vec{w}' e^{-ik \vec{g}^T \vec{w}'} e^{-\frac{1}{2} \vec{w}'^T S_N^{-1} \vec{w}'} \quad \text{⊖}$$

⊖ M is symm. & invertible,
Hermitian conjugate

$$\left[\begin{aligned} \vec{x}^+ M \vec{x} - 2 \vec{b}^+ \vec{x} &= (\vec{x} - M^{-1} \vec{b})^+ M (\vec{x} - M^{-1} \vec{b}) - \\ &- \vec{b}^+ M^{-1} \vec{b} \end{aligned} \right]$$

Here, $\begin{cases} \vec{x} \rightarrow \vec{w}', \\ M \rightarrow \frac{S_N^{-1}}{2} \Rightarrow M^{-1} = 2 S_N, \\ 2 \vec{b}^+ \rightarrow ik \vec{g}^T \Rightarrow \vec{b} \rightarrow -\frac{ik}{2} \vec{g}. \end{cases}$

$$\text{⊖} \quad \frac{1}{(2\pi)^{M/2}} \frac{1}{|S_N|^{1/2}} \frac{1}{2\pi} \int dk e^{ik(a - \vec{g}^T \vec{w}_{MAP})_x} e^{-\frac{k^2}{4} \vec{g}^T (2 S_N) \vec{g}_x} \\ \times \int d\vec{w}' e^{-\underbrace{(\vec{w}' + ik S_N \vec{g})^+ \frac{S_N^{-1}}{2} (\vec{w}' + ik S_N \vec{g})}_{(2\pi)^{\frac{M}{2}} |S_N|^{1/2}}} = \\ = \frac{1}{2\pi} \int dk e^{ik(a - \vec{g}^T \vec{w}_{MAP})_x} e^{-\frac{k^2}{2} \vec{g}^T S_N \vec{g}_x}$$

Using $\int dx e^{ax^2+bx} = \sqrt{-\frac{\pi}{a}} e^{-b^2/4a}$,
 we obtain:

$$\begin{cases} x \rightarrow k, \\ a \rightarrow -\frac{1}{2} \bar{y}^T S_N \bar{y}, \\ b \rightarrow i(a - \bar{y}^T \bar{w}_{MAP}) \end{cases}, \text{ so that}$$

$$p(a) = \frac{1}{2\pi} \underbrace{\sqrt{\frac{2\pi}{\bar{y}^T S_N \bar{y}}}}_{\frac{1}{\sqrt{2\pi\sigma_a^2}}} e^{-\frac{(a - \bar{y}^T \bar{w}_{MAP})^2}{2 \underbrace{\bar{y}^T S_N \bar{y}}_{\sigma_a^2}}} = \mathcal{N}(a | \mu_a, \sigma_a^2)$$

Thus

$$p(c_1 | \bar{y}, \bar{E}) = \int da \underbrace{\sigma(a)}_{\text{use } \tilde{\varphi}(\lambda a)} \mathcal{N}(a | \mu_a, \sigma_a^2) x =$$

$$\tilde{\varphi}\left(\frac{\mu_a}{(\lambda^2 + \sigma_a^2)^{1/2}}\right) = \sigma\left(\frac{\mu_a}{(1 + \lambda^2 \underbrace{\sigma_a^2}_{\frac{\pi}{8}})^{1/2}}\right)$$

shown
below

Note that the DB ($p(c_1 | \bar{y}, \bar{E}) = p(c_2 | \bar{y}, \bar{E}) = 0.5$)
 is given by $\mu_a = 0$,
 $\bar{w}_{MAP}^T \bar{y}$, linear in feature space

Now, let's show that

$$\int da \phi(\lambda a) \mathcal{N}(a | \mu, \sigma^2) = \phi\left(\frac{\mu}{\sqrt{\frac{1}{\lambda^2} + \sigma^2}}\right)$$

RHS:

$$\phi\left(\frac{\mu}{\sqrt{\frac{1}{\lambda^2} + \sigma^2}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z d\theta e^{-\theta^2/2}, \quad \text{s.t.}$$

$$\frac{d\phi}{d\mu} = \frac{d\phi}{dz} \frac{dz}{d\mu} = \frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2(\frac{1}{\lambda^2} + \sigma^2)}} \frac{1}{\sqrt{\frac{1}{\lambda^2} + \sigma^2}}$$

LHS:

$$\text{use } \frac{a-\mu}{\sigma} = z \Rightarrow a = \mu + \sigma z$$

$$da = \sigma dz$$

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dz \underbrace{\phi(\lambda(\mu + \sigma z))}_{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda(\mu + \sigma z)} d\theta e^{-\theta^2/2}} e^{-z^2/2}$$

$$\text{Then } \frac{d}{d\mu} (\text{LHS}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dz e^{-\frac{\lambda^2(\mu + \sigma z)^2}{2}} \lambda e^{-z^2/2} =$$

$$= \frac{\lambda}{2\pi} e^{-\frac{\lambda^2\mu^2}{2}} \int dz e^{-\lambda^2\mu\sigma z} e^{-\frac{z^2}{2}(1 + \lambda^2\sigma^2)} \quad \textcircled{=}$$

$$\text{Here, } \begin{cases} a = -\frac{1}{2}(1 + \lambda^2\sigma^2) \\ b = -\lambda^2\mu\sigma \end{cases}$$

$$\textcircled{=} \frac{\lambda}{2\pi} e^{-\frac{\lambda^2\mu^2}{2}} \sqrt{\frac{2\pi}{1 + \lambda^2\sigma^2}} e^{\frac{\lambda^4\mu^2\sigma^2}{2(1 + \lambda^2\sigma^2)}} \quad \textcircled{=}$$

$$\Leftrightarrow \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\lambda^{-2} + \zeta^2}} e^{-\frac{\lambda^2 \mu^2}{2} \left[1 - \frac{\lambda^2 \zeta^2}{1 + \lambda^2 \zeta^2} \right]}$$

$$\frac{\lambda^2 \mu^2}{2} \frac{1}{1 + \lambda^2 \zeta^2} = \frac{\mu^2}{2} \frac{1}{\lambda^{-2} + \zeta^2}$$

So, $\frac{d}{d\mu}(\text{LHS}) = \frac{d}{d\mu}(\text{RHS}) \Rightarrow \text{LHS} = \text{RHS} + \text{const}(\mu)$

Consider the $\mu \rightarrow +\infty$ limit:

$$\lim_{\mu \rightarrow \infty} \text{RHS} = 1$$

$$\lim_{\mu \rightarrow \infty} \text{LHS} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dz e^{-\frac{z^2}{2}} = 1$$

$$\phi(\lambda(\mu + \zeta z)) \rightarrow 1 \text{ as } \mu \rightarrow +\infty$$

Thus $\text{const}(\mu) = 0 \Rightarrow \underline{\underline{\text{LHS} = \text{RHS}}}$