office hour 3:30pm
KO2 Kavli

Methods for computing Z

1. Analytically evaluating integral (conjugate prior)

param X we Gaussian prior $\pi(X) = \dfrac{\exp\left(-\frac{1}{2}\frac{(X-\mu)^2}{\Sigma^2}\right)}{\sqrt{2\pi}\,\Sigma}$

data $\{x_i\}$ N measurement, independent $x_i \sim N(X, \sigma^2)$

$$\mathcal{L}(\{x_i\} \mid X) = \prod_{i=1}^{N} \frac{\exp\left(-\frac{1}{2}\frac{(X-x_i)^2}{\sigma^2}\right)}{\sqrt{2\pi}\,\sigma}$$

Bayes' $\Rightarrow$ $P(X \mid \{x_i\}) = \frac{1}{Z} \mathcal{L}(\{x_i\} \mid X)\, \pi(X)$

$$P(X|\{x_i\}) = Z^{-1}(2\pi)^{-\left(\frac{N+1}{2}\right)}\sigma^{-N}\Sigma^{-1}\exp\left(\frac{-1}{2}\left[\frac{\sum_i (X-x_i)^2}{\sigma^2} + \frac{(X-\mu)^2}{\Sigma^2}\right]\right)$$

$$= \underline{\hspace{4cm}}\ \exp\left(\frac{-1}{2}\left[\frac{N\Sigma^2+\sigma^2}{\sigma^2\Sigma^2}X^2 - 2X\left(\frac{\sum_i x_i}{\sigma^2} + \frac{\mu}{\Sigma^2}\right) + \left(\frac{\sum_i x_i^2}{\sigma^2} + \frac{\mu^2}{\Sigma^2}\right)\right]\right)$$

$$= \underline{\hspace{4cm}}\ \exp\left(\frac{-1}{2}\boxed{\frac{N\Sigma^2+\sigma^2}{\sigma^2\Sigma^2}}\left[X^2 - 2X\boxed{\frac{\Sigma^2\sum_i x_i + \sigma^2\mu}{N\Sigma^2+\sigma^2}} + \frac{\Sigma\sum_i x_i^2 + \sigma^2\mu^2}{N\Sigma^2+\sigma^2}\right]\right)$$

$\boxed{1}$

$$= Z^{-1}(2\pi)^{-\left(\frac{N+1}{2}\right)}\sigma^{-N}\Sigma^{-1}\exp\left(-\frac{1}{2}\frac{(X-\mu')^2}{\Sigma'^2} + \frac{1}{2}\frac{\mu'^2}{\Sigma'^2} - \frac{1}{2}\frac{\mu^2}{\Sigma^2} - \frac{\sum_i x_i^2}{2\sigma^2}\right)$$

$$\boxed{\mu' = \frac{\Sigma^2\sum_i x_i + \sigma^2\mu}{N\Sigma^2+\sigma^2}} \qquad \boxed{\Sigma'^2 = \frac{\sigma^2\Sigma^2}{N\Sigma^2+\sigma^2}}$$

we recognise $\quad P(X \mid \{x_i\}) = \dfrac{\exp\left(-\frac{1}{2}\frac{(X-\mu')^2}{\Sigma'^2}\right)}{\sqrt{2\pi}\,\Sigma'}$ $\quad \boxed{2}$

so equate $\boxed{1}$ & $\boxed{2}$

$$\underbrace{Z = (2\pi)^{-N/2}\sigma^{-N}\frac{\Sigma'}{\Sigma}\exp\left(-\frac{1}{2}\left[\frac{\mu^2}{\Sigma^2} - \frac{\mu'^2}{\Sigma'^2} + \frac{\sum_i x_i^2}{\sigma^2}\right]\right)}$$

answer to part vi on WS1

μ', Σ' to be understood as functions of $\mu, \Sigma, N, \sigma, \{x_i\}$

conjugate priors gives the evidence for free (without need for integration).

<u>Laplace's approximation</u>

$$Z = \int_x dx \underbrace{\mathcal{L}(d|x)\,\pi(x)}_{\substack{\text{unnormalised posterior} \\ P^*(x|d)}}$$

e.g. in 1 dimension

$$\log P^*(x|d) \approx \log P^*(\underset{\substack{\uparrow \\ \text{the peak of the unnormalised posterior}}}{\hat{x}}|d) - \frac{1}{2}(x-\hat{x})^2 c + O((x-\hat{x})^3)$$

where $c = - \frac{d^2}{dx^2}\Big|_{x=\hat{x}} \log P^*(x|d)$

within this approximation $P^*(x|d) \sim$ Gaussian shape.

As a consequence of CLT, this will eventually be a good approximation in the limit of large sample or large SNR.

$$P^*(x|d) \approx P^*(\hat{x}|d) \exp\left(-\frac{1}{2}(x-\hat{x})^2 c\right)$$

$$\Rightarrow Z = \int dx\, P^*(x|d) = \underline{P^*(\hat{x}|d)\sqrt{\frac{2\pi}{c}}}$$

Laplace approximation in 1D

recall $Z = \int dx \; P^*(x \mid \alpha)$

change of params $x \to y(x)$

LA involves $c = -\frac{d^2}{dx^2}\Big|_{x=\hat{x}} \log P^*(x \mid \alpha)$

LA to the Z depends on the parameters used.

LA in d dimensions . param vec $\vec{x} \in \mathbb{R}^D$

$$\log P^*(\vec{x} \mid d) = \log \mathcal{L}(\cancel{d} \mid \vec{x}) + \log \pi(\vec{x})$$

found peak $\hat{\vec{x}}$, multivariate Taylor series

$$\log P^*(\vec{x} \mid d) = \log P^*(\hat{\vec{x}} \mid d) \cancel{+} - \frac{1}{2} \sum_{ij} C_{ij} (x_i - \hat{x}_i)(x_j - \hat{x}_j) + \mathcal{O}(x^3)$$

where $C_{ij} = - \frac{\partial^2}{\partial x_i \partial x_j} \Big|_{\vec{x} = \hat{\vec{x}}} \log P^*(\vec{x} \mid d)$

Hessian matrix .

$$Z = P^*(\hat{\vec{x}} \mid d) \sqrt{\frac{(2\pi)^D}{\det \underline{\underline{C}}}}$$

LA in D dimensions .

if $P^*(\vec{x} \mid d)$ has more than 1 peak.

use LA to integrate under peak, and then sum results.

(we are assuming peaks ~~well~~ far apart in parameter space)