3. Probabilistic Inference

Mittwoch, 6. November 2019 13:15

Consider the following probabilistic model

$$p(\mu \mid \alpha) = \mathcal{N}(\mu \mid 0, \alpha^{-1}) = \sqrt{\frac{\alpha}{2\pi}} \exp\left(-\frac{\alpha}{2}\mu^2\right) \quad \text{- prior}$$

Note that here we don't parametrize the variance σ^2, but rather specify the *precision* parameter $\alpha = 1/\sigma^2$.

$$p(x \mid \mu) = \mathcal{N}(x \mid \mu, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right) \quad \text{- likelihood}$$

You are given a set of observations $\mathcal{D} = \{x_1, ..., x_N\}$ consisting of N samples $x_i \in \mathbb{R}$.

$$\mathcal{N}(\mu, \sigma^2)$$

$$\alpha = \frac{1}{\sigma^2}$$

Problem 6: Derive the maximum likelihood estimate μ_{MLE}. Show your work.

$$\mu_{MLE} = \underset{\mu}{\arg\max}\ p(D \mid \mu)$$

$$= \underset{\mu}{\arg\max}\ \log p(D \mid \mu)$$

1. Compute / find the expr. for $\log p(D \mid \mu)$

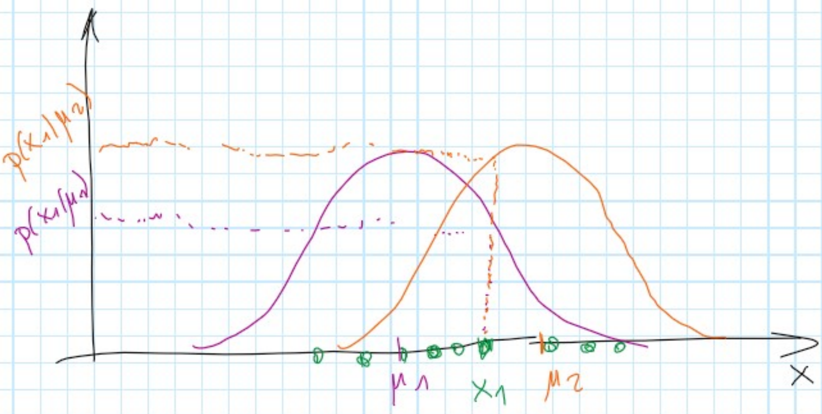
2. $\frac{\partial}{\partial \mu} \log p(D \mid \mu) \overset{!}{=} 0$ solve for μ

$$\log p(D \mid \mu) = \log p(x_1, \dots, x_N \mid \mu)$$

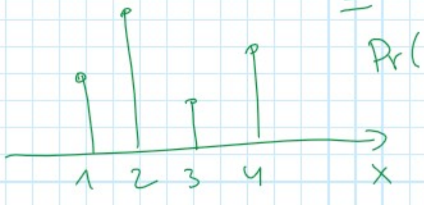
$$= \log\left(\prod_{i=1}^{N} p(x_i \mid \mu)\right)$$

$$= \sum_{i=1}^{N} \log p(x_i \mid \mu)$$

$$\log(a \cdot b) = \log a + \log b$$

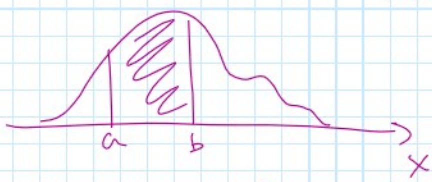


PMF



$$\Pr(X=2) = p(2)$$

PDF



$$\Pr(X \in [a,b]) = \int_a^b p(x)\, dx$$

$$= \sum_{i=1}^{N} \log \mathcal{N}(x_i \mid \mu, 1)$$

$$= \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)\right)$$

$$= \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}(x_i-\mu)^2\right)\right)$$

$$= \sum_{i}\left[\underbrace{\log\left(\frac{1}{\sqrt{2\pi}}\right)}_{\text{const. w.r.t. } \mu} + \log\left(\exp\left(-\frac{1}{2}(x_i-\mu)^2\right)\right)\right]$$

$$= \sum_{i}\left(-\frac{1}{2}(x_i-\mu)^2\right) + \text{const.}$$

$$= -\frac{1}{2}\sum_{i}\left(x_i^2 - 2x_i\mu + \mu^2\right) + \text{const.}$$

$$= \underbrace{-\frac{1}{2}\sum_{i} x_i^2}_{\text{const.}} + \frac{2}{2} \cdot \sum_{i} x_i\mu - \frac{1}{2}\sum_{i=1}^{N}\mu^2 \quad \underline{+ \text{const.}}$$

$$\log p(D|\mu) = \mu \cdot \sum_{i} x_i - \frac{1}{2} N \cdot \mu^2 + \text{const.}$$

Recall, that we want to maximize $\log p(D|\mu)$, so we solve $\frac{\partial}{\partial\mu} \log p(D|\mu) \overset{!}{=} 0$

$$\frac{\partial}{\partial\mu} \log p(D|\mu) = \sum_{i=1}^{N} x_i - \frac{N}{2} \cdot 2\mu$$

$$= -N \cdot \mu + \sum_{i=1}^{N} x_i \overset{!}{=} 0$$

$$-N \cdot \mu + \sum_{i} x_i = 0$$

$$\sum_{i=1}^{N} x_i = N \cdot \mu$$

$$\Rightarrow \quad \mu = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i$$

$\Rightarrow$ μ_{MLE} is the <u>average</u> of the data points.

$\Rightarrow \mu_{MLE}$ is the average of the data points.

Problem 7: Derive the maximum a posteriori estimate μ_{MAP}. Show your work.

$$\mu_{MAP} = \underset{\mu}{\operatorname{argmax}}\ p(\mu|D)$$

$$= \underset{\mu}{\operatorname{argmax}}\ \log p(\mu|D)$$

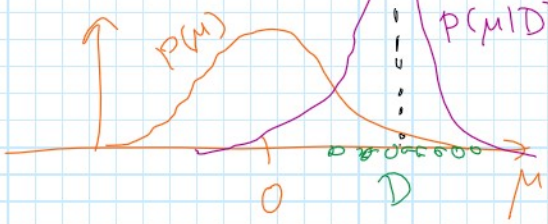
$$= \underset{\mu}{\operatorname{argmax}}\ \log\left(\frac{p(D|\mu)\cdot p(\mu)}{p(D)}\right)$$

$$= \underset{\mu}{\operatorname{argmax}}\left(\log p(D|\mu) + \log p(\mu) - \underbrace{\log p(D)}_{\text{const.}}\right)$$

$$= \underset{\mu}{\operatorname{argmax}}\left[\underbrace{\log p(D|\mu)}_{\text{we know this from before}} + \log p(\mu)\right]$$

$p(\mu)$ - initial beliefs (prior)

$p(\mu|D)$ - updated beliefs (posterior)



In MLE

$$\underset{\mu}{\operatorname{argmax}}\ \log p(D|\mu)$$

$$\log p(D|\mu) = \mu \cdot \sum_{i=1}^{N} x_i - \frac{N}{2}\cdot\mu^2$$

$$\log p(\mu|\alpha) = \log \mathcal{N}(\mu\,|\,0, \alpha^{-1})$$

$$= \log\left(\sqrt{\frac{\alpha}{2\pi}}\cdot \exp\left(-\frac{\alpha}{2}\cdot(\mu-0)^2\right)\right)$$

$$= \underbrace{\log\left(\sqrt{\frac{\alpha}{2\pi}}\right)}_{\text{const.}} + \log\left(\exp\left(-\frac{\alpha}{2}\mu^2\right)\right)$$

$$= -\frac{\alpha}{2}\cdot\mu^2 + \text{const.}$$

$$\log p(\mu|D) = \log p(D|\mu) + \log p(\mu|\alpha) + \text{const.}$$

$$= \mu\cdot\sum_{i=1}^{N} x_i - \frac{N}{2}\mu^2 - \frac{\alpha}{2}\mu^2 + \text{const.}$$

$$\frac{\partial}{\partial\mu}\log p(\mu|D) = \sum_i x_i - N\cdot\mu - \alpha\cdot\mu \overset{!}{=} 0$$

$$\frac{\partial}{\partial \mu} \log p(\mu \mid D) = \sum_i x_i - N \cdot \mu - \alpha \cdot \mu \stackrel{!}{=} 0$$

$$\Leftrightarrow \quad \sum_{i=1}^{N} x_i = (N + \alpha) \cdot \mu$$

Recall,

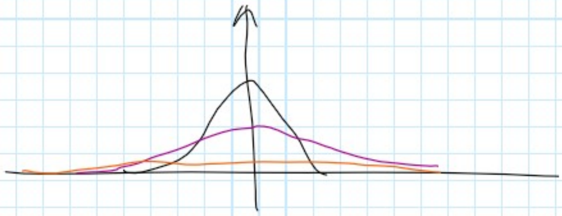
$$\Leftrightarrow \quad \boxed{\mu = \frac{1}{N+\alpha} \cdot \sum_{i=1}^{N} x_i} \quad \mu_{MAP} \qquad \mu_{MLE} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i$$

Problem 8: Does there exist a prior distribution over μ such that $\mu_{\text{MLE}} = \mu_{\text{MAP}}$? Justify your answer.

If we could set $\alpha = 0$, we would have

$$\mu_{MLE} = \mu_{MAP}$$

$$\alpha = \frac{1}{\sigma^2} \qquad \alpha = 0 \Leftrightarrow \sigma^2 = \infty$$



$$\int_{-\infty}^{\infty} c \, d\mu = \infty$$

$$p(\mu) \propto c$$

$$\log p(\mu) = \log c$$

$$\arg\max_{\mu} \; \log p(D \mid \mu) + \log p(\mu)$$

$$= \arg\max_{\mu} \; \log p(D \mid \mu) + \underline{\underline{\log c}}$$

Problem 9: Derive the posterior distribution $p(\mu \mid \mathcal{D}, \alpha)$. Show your work.

$$\boxed{p(\mu \mid D, \alpha)} = \frac{p(D \mid \mu) \cdot p(\mu \mid \alpha)}{p(D \mid \alpha)} \qquad p(D \mid \alpha) = \int p(D \mid \mu, \alpha) \cdot p(\mu \mid \alpha) \, d\mu$$

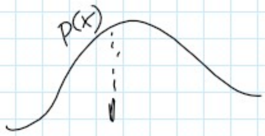
$$= \int p(D \mid \mu) \, p(\mu \mid \alpha) \, d\mu$$

$$\underline{\underline{\propto}} \; p(D \mid \mu) \cdot p(\mu \mid \alpha)$$

$$\propto \left(\prod_{i=1}^{N} p(x_i \mid \mu) \right) \cdot p(\mu \mid \alpha)$$

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \propto \left(\prod_{i=1}^{N} \mathcal{N}(x_i \mid \mu, 1) \right) \cdot \mathcal{N}(\mu \mid 0, \alpha^{-1})$$

$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \propto \left(\prod_{i=1} \mathcal{N}(x_i \mid \mu, 1)\right) \cdot \mathcal{N}(\mu \mid 0, \alpha^{-1})$



$$\propto \left(\prod_i \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}(x_i-\mu)^2\right)\right) \cdot \sqrt{\frac{\alpha}{2\pi}} \cdot \exp\left(-\frac{\alpha}{2}\mu^2\right)$$

const   const.

$$\propto \left(\prod_{i=1}^{N} \exp\left(-\frac{1}{2}(x_i-\mu)^2\right)\right) \cdot \exp\left(-\frac{\alpha}{2}\mu^2\right)$$

$\exp(a) \cdot \exp(b)$
$= \exp(a+b)$

$$\propto \exp\left(\sum_{i=1}^{N}\left(-\frac{1}{2}(x_i-\mu)^2\right)\right) \cdot \exp\left(-\frac{\alpha}{2}\mu^2\right)$$

$$\propto \exp\left(\left[-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right] - \frac{\alpha}{2}\mu^2\right)$$

$$\propto \exp\left(\underbrace{-\frac{1}{2}\sum_{i=1}^{N} x_i^2}_{\text{const.}} + \mu \cdot \sum_{i=1}^{N} x_i - \frac{1}{2}\sum_{i=1}^{N}\mu^2 - \frac{\alpha}{2}\mu^2\right)$$

$$\propto \exp\left(\mu \cdot \sum_{i=1}^{N} x_i - \frac{N}{2}\mu^2 - \frac{\alpha}{2}\mu^2\right)$$

$$\propto \exp\left(-\frac{N+\alpha}{2}\mu^2 + \mu \sum_{i=1}^{N} x_i\right)$$

$$\mathcal{N}(\mu \mid m, \beta^{-1}) = \sqrt{\frac{\beta}{2\pi}} \cdot \exp\left(-\frac{\beta}{2}(\mu-m)^2\right)$$

$$= \sqrt{\frac{\beta}{2\pi}} \cdot \exp\left(-\frac{\beta}{2}\mu^2 + \beta \cdot m \cdot \mu - \frac{\beta}{2}m^2\right)$$

$$= \underbrace{\sqrt{\frac{\beta}{2\pi}} \cdot \exp\left(-\frac{\beta}{2}m^2\right)} \cdot \underbrace{\exp\left(-\frac{\beta}{2}\mu^2 + \beta m \cdot \mu\right)}$$

$$= \underbrace{\sqrt{\frac{\beta}{2\pi}} \cdot \exp\left(-\frac{\beta}{2} m^2\right)}_{\text{const. w.r.t. } \mu} \cdot \exp\left(-\frac{\beta}{2}\mu^2 + \beta m \cdot \mu\right)$$

We want to find $\underline{m, \beta}$ s.t. our posterior

$$p(\mu | D, \alpha) = \mathcal{N}(\mu | m, \beta^{-1})$$

$$-\frac{N+\alpha}{2} \cdot \mu^2 \overset{!}{=} -\frac{\beta}{2}\mu^2$$

$$\beta = N + \alpha$$

$$\text{var} = \frac{1}{\text{precision}}$$

$$\mu \cdot \sum_i x_i \overset{!}{=} \beta \cdot m \cdot \mu$$

$$\beta = \frac{1}{\text{variance}}$$

$$\beta \cdot m = \sum_i x_i$$

$$(N+\alpha) \cdot m = \sum_i x_i$$

$$m = \frac{1}{N+\alpha} \cdot \sum_i x_i$$

$$p(\mu | D, \alpha) = \mathcal{N}\Big(\mu \,\Big|\, \overbrace{\frac{1}{N+\alpha} \cdot \sum_i x_i}^{\text{mean}}, \overbrace{(N+\alpha)^{-1}}^{\text{precision}}\Big)$$

$$\mu_{MAP} = \frac{1}{N+\alpha} \sum_i x_i$$

Problem 10: Derive the posterior predictive distribution $p(x_{\cancel{new}} \mid \mathcal{D}, \alpha)$. Show your work.

We know $p(\mu | D, \alpha)$

$$p(x \mid D, \alpha) = \int p(x \mid \mu, D, \alpha) \cdot p(\mu \mid D, \alpha)\, d\mu$$

$$= \int p(x \mid \mu) \cdot p(\mu \mid D, \alpha)\, d\mu$$

$$p(\mu) = \sqrt{\frac{\alpha}{2\pi}} \exp\left(-\frac{\mu^2}{2}\right)$$

$$= (x - \mu - 0)^2$$

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2}\right)$$

$$= \int \mathcal{N}(x \mid \mu, 1) \cdot \mathcal{N}(\mu \mid m, \beta^{-1})\, d\mu$$

$$\int \mathcal{N}(x - \mu \mid 0, 1) \cdot \mathcal{N}(\mu \mid m, \beta^{-1})\, d\mu$$

$$\int p(x \mid \mu) \cdot p(\mu \mid D, \alpha)\, d\mu$$

$$p(x \mid D, \alpha)$$

$$= \mathbb{E}_{\mu \sim p(\mu \mid D, \alpha)}\left[p(x \mid \mu)\right]$$

1. $\mu \sim p(\mu \mid D, \alpha)$
2. $x \sim p(x \mid \mu)$

1

$y \sim \mathcal{N}(y \mid c, \sigma^2)$

$x \sim \mathcal{N}(x \mid d, \tau^2)$

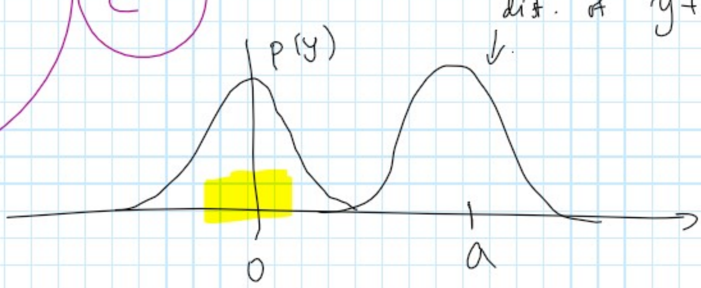
$x + y \sim \mathcal{N}(x + y \mid c + d, \sigma^2 + \tau^2)$

Alternatively,

1. $\mu \sim p(\mu \mid D, \alpha)$
2. $y \sim \mathcal{N}(y \mid 0, 1)$
3. $x = y + \mu$

2

$y + a \sim$



$\text{var} = \frac{1}{\beta}$

$\mu \sim \mathcal{N}(\mu \mid m, \beta^{-1})$

$\text{var} = 1$

$y \sim \mathcal{N}(y \mid 0, 1)$

$x = y + \mu$

$\frac{1}{\beta}$

$$\frac{1}{\left(1 + \frac{1}{\beta}\right)}$$

$$\frac{1}{\beta} + 1$$

$X = y + \mu$

$x \sim \mathcal{N}(x \mid m + 0,\ \beta^{-1} + 1)$