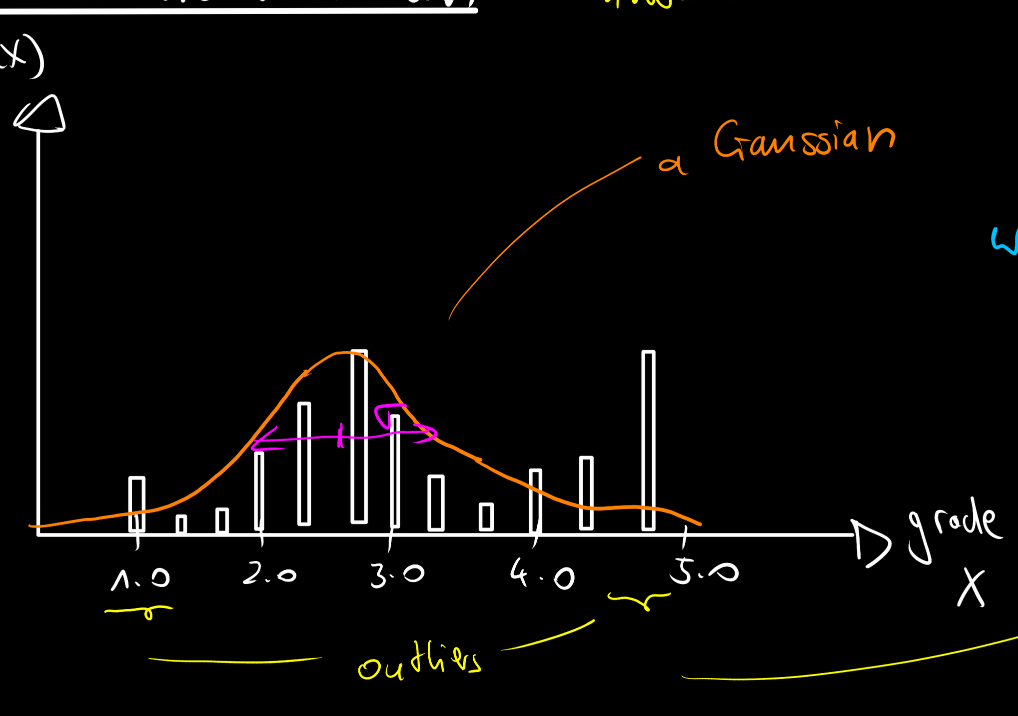
Posterior and MAP for the Gaussian/Normal with unknown mean

univariate



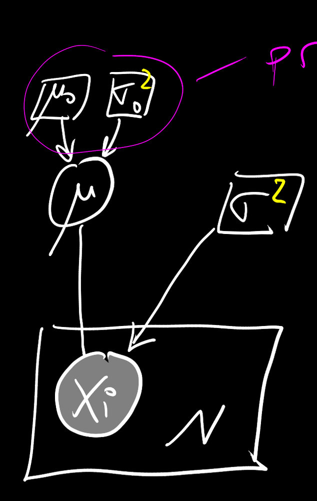
fair teacher
fix σ (or σ^2)

where is mu/mean?

→ MLE

robuster?
→ prior knowledge

DGM



now we have dataset

$$D = \underbrace{\{2.2,\ 1.7,\ 2.7,\ \ldots\}}_{N \text{ grades}}$$

likelihood:

$$\mathcal{L}(D \mid \mu; \sigma^2) \overset{i.i.d.}{=} \prod_{i=0}^{N-1} p(X_i \mid \mu)$$

$$= \prod_{i=0}^{N-1} \mathcal{N}(X_i \mid \mu; \sigma^2)$$

$$= \prod_{i=0}^{N-1} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} \cdot (X_i - \mu)^2\right)$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left(-\frac{1}{2\sigma^2} \cdot \sum_{i=0}^{N-1} (X_i - \mu)^2\right)$$

μ appears in exp squared
↳ use Normal/Gaussian as prior

$$\mu \sim \mathcal{N}(\mu; \mu_0, \sigma_0^2)$$

$$X_i \sim \mathcal{N}(X_i \mid \mu; \sigma^2)$$

joint distribution

$$p(\mu, \underline{X} = D) = p(\mu)\, p(\underline{X} = D \mid \mu)$$

$p(\mu, D)$

$$= p(\mu) \prod_{i=0}^{N-1} p(X_i = x^{[i]} \mid \mu)$$

$$= \mathcal{N}(\mu; \mu_0, \sigma_0^2) \prod_{i=0}^{N-1} \mathcal{N}(X_i = x^{[i]} \mid \mu; \sigma^2)$$

$$= \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left(-\frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2\right)$$

$$= \frac{1}{\sigma_0\sqrt{2\pi}} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2 - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2\right)$$

posterior

$$p(\mu \mid D) \overset{\text{Bayes' Rule}}{=} \frac{p(D \mid \mu)\, p(\mu)}{p(D)}$$

$$\sim p(D \mid \mu)\, p(\mu) = p(\mu, D)$$

→ conditional is proportional to the joint

$$p(\mu \mid D) \sim \cancel{\frac{1}{\sigma_0\sqrt{2\pi}}} \cancel{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2 - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (x^{[i]} - \mu)^2\right)$$

$$\sim \exp\left(-\frac{1}{2\sigma_0^2}(\mu^2 - 2\mu\mu_0 + \mu_0^2) - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} ({x^{[i]}}^2 - 2x^{[i]}\mu + \mu^2)\right)$$

$$= \exp\left(-\frac{\mu^2}{2\sigma_0^2} + \frac{\mu\mu_0}{\sigma_0^2} - \frac{\mu_0^2}{2\sigma_0^2} - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (x^{[i]})^2 + \frac{\mu}{\sigma^2} \sum_{i=0}^{N-1} (x^{[i]}) - \frac{\mu^2}{2\sigma^2} \cdot N\right)$$

$$= \exp\left(-\frac{\mu^2}{2\sigma_0^2} - \frac{\mu^2}{2\sigma^2} N + \frac{\mu\mu_0}{\sigma_0^2} + \frac{\mu}{\sigma^2} \sum_{i=0}^{N-1} (x^{[i]}) \cancel{- \frac{\mu_0^2}{2\sigma_0^2}} \cancel{- \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (x^{[i]})^2}\right)$$

$$\sim \exp\left(-\frac{\mu^2\sigma^2 + \mu^2\sigma_0^2 N}{2\sigma_0^2\sigma^2} + \frac{\mu\mu_0\sigma^2 + \mu\sigma_0^2 \sum_{i=0}^{N-1} (x^{[i]})}{\sigma_0^2\sigma^2}\right)$$

$$= \exp\left(-\frac{1}{2\sigma_0^2\sigma^2}\left(\underbrace{(\sigma^2 + N\sigma_0^2)}_{a}\mu^2 - \underbrace{2\left(\mu_0\sigma^2 + \sigma_0^2 \sum_{i=0}^{N-1} (x^{[i]})\right)}_{b}\mu \underbrace{+\, 0}_{c}\right)\right)$$

$$= \exp\left(-\frac{1}{2\sigma_0^2\sigma^2}(a\mu^2 + b\mu + c)\right)$$

$$= \exp\left(-\frac{1}{2\sigma_0^2\sigma^2}(a(\mu - d)^2 + e)\right)$$

Completing the square

$a = a \qquad d = -\frac{b}{2a} \qquad e = c - \frac{b^2}{4a}$

$$d = \cancel{-}\frac{\cancel{-}\cancel{2}(\sigma^2\mu_0 + \sigma_0^2 \sum_{i=0}^{N-1} x^{[i]})}{\cancel{2} \cdot (\sigma^2 + N\sigma_0^2)}$$

$$e = 0 - \frac{(\cancel{-}\cancel{2}(\sigma^2\mu_0 + \sigma_0^2 \sum_{i=0}^{N-1} x^{[i]}))^2}{\cancel{4} \cdot (\sigma^2 + N\sigma_0^2)}$$

$$p(\mu \mid D) \sim \exp\left(-\frac{1}{2\sigma_0^2\sigma^2} \cdot \left((\sigma^2 + N\sigma_0^2) \cdot \left(\mu - \frac{\sigma^2\mu_0 + \sigma_0^2 \sum_{i=0}^{N-1} x^{[i]}}{\sigma^2 + N\sigma_0^2}\right)^2 \cancel{- \frac{\sigma^2\mu_0 + \sigma_0^2 \sum_{i=0}^{N-1} x^{[i]}}{\sigma^2 + N\sigma_0^2}}\right)\right)$$

$$\sim \exp\left(-\frac{\sigma^2 + N\sigma_0^2}{2\sigma_0^2\sigma^2}\left(\mu - \underbrace{\frac{\sigma^2\mu_0 + \sigma_0^2 \sum_{i=0}^{N-1} x^{[i]}}{\sigma^2 + N\sigma_0^2}}_{\mu_N}\right)^2\right)$$

isn't it just another Normal

$$\cancel{-}\frac{1}{\cancel{2}\sigma_N^2} = \cancel{-}\frac{\sigma^2 + N\sigma_0^2}{\cancel{2}\sigma_0^2\sigma^2}$$

$$\sigma_N = \sqrt{\frac{\sigma_0^2\sigma^2}{\sigma^2 + N\sigma_0^2}}$$

$$p(\mu \mid D) = \mathcal{N}(\mu; \mu_N, \sigma_N^2)$$

$$\boxed{\mu_N = \frac{\sigma^2\mu_0 + \sigma_0^2 \sum_{i=0}^{N-1} x^{[i]}}{\sigma^2 + N\sigma_0^2}}$$

$$\boxed{\mu_{MLE} = \frac{1}{N} \sum_{i=0}^{N-1} x^{[i]}}$$

$$= \frac{\sigma^2}{\sigma^2 + N\sigma_0^2}\mu_0 + \frac{\sigma_0^2 N}{\sigma^2 + N\sigma_0^2}\mu_{MLE}$$

→ a linear combination (= a weighted average) of MLE and prior knowledge

$$\boxed{\sigma_N = \frac{\sigma_0\sigma}{\sqrt{\sigma^2 + N\sigma_0^2}}}$$

$$\frac{1}{\sigma_N^2} = \frac{\sigma^2 + N\sigma_0^2}{\sigma_0^2\sigma^2}$$

$$= \frac{\cancel{\sigma^2}}{\sigma_0^2\cancel{\sigma^2}} + \frac{N\cancel{\sigma_0^2}}{\cancel{\sigma_0^2}\sigma^2}$$

$$= \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$

sth like the harmonic mean

What is the MAP?

↳ Mode of the posterior

$$\rightarrow \boxed{\mu_{MAP} = \mu_N = \frac{\sigma^2\mu_0 + \sigma_0^2 \sum_{i=0}^{N-1} x^{[i]}}{\sigma^2 + N\sigma_0^2}}$$