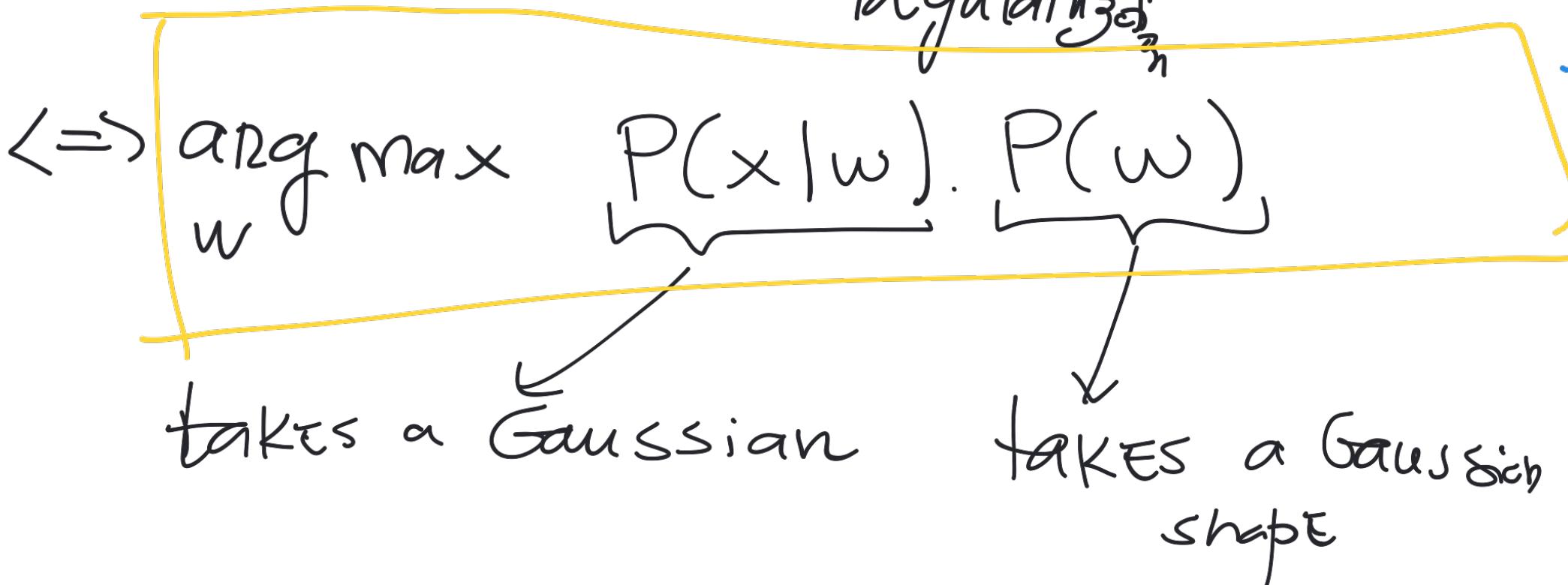


MAP

$$\arg \min_w \frac{1}{2} \sum_{i=1}^N (t_i - y_i)^2 + \lambda \cdot \sum_{j=0}^n w_j^2$$

Ridge
Regularized



- When we use Lasso regularizer:
the prior $P(w)$ is Laplacian distribution

Example: Experiment flipping a coin 3 times and observing the results H H H.

- Hidden state: "what coin was used?"
(fair, 2-headed, 2-tailed)

$$\begin{array}{c} \mu \equiv \text{prob. of heads} \\ \hline H \equiv 1, T \equiv 0, S = \{0, 1\} \end{array}$$

$$P(xe = 1 | \mu) = \mu$$

$$P(xe = 0 | \mu) = 1 - \mu$$

Data Likelihood:

$$P(x|\mu) = \mu^x \cdot (1-\mu)^{1-x}$$

$$\text{if } x=0: P(x|\mu) = \mu^0 \cdot (1-\mu)^1 = 1-\mu$$

Bernoulli distribution

Now we have $\{x_i\}_{i=1}^N$, assume that each sample x_i is independent and are identically distributed (i.i.d.)

$E \equiv$ Event of flip coin N times

$$P(E|\mu) = P(x_1, x_2, \dots, x_N | \mu)$$

$$= P(x_1|\mu) \cdot P(x_2|\mu) \cdot \dots \cdot P(x_N|\mu)$$

$$\xrightarrow{\text{indep.}} = \prod_{i=1}^n P(x_i|\mu)$$

$$\xrightarrow{\text{identically dist.}} = \prod_{i=1}^n \mu^{x_i} \cdot (1-\mu)^{1-x_i}$$

Binomial
data likelihood

MLE

$$\arg \max_{\mu} P(E|e)$$

$$\Leftrightarrow \arg \max_{\mu} \ln(P(E|\mu))$$

↑
ln is a monotonic fct.

$$L = \ln(P(E|\mu)) = \sum_{i=1}^N x_i \cdot \ln(\mu) + (1-x_i) \cdot \ln(1-\mu)$$

$$\frac{\partial L}{\partial \mu} = 0 \Leftrightarrow \sum_{i=1}^N x_i \cdot \frac{1}{\mu} + (1-x_i) \cdot \left(-\frac{1}{1-\mu}\right) = 0$$

$$\Leftrightarrow \frac{1}{\mu} \sum_{i=1}^N x_i - \frac{1}{1-\mu} \cdot \sum_{i=1}^N (1-x_i) = 0$$

$$\Leftrightarrow \frac{1}{\mu} \sum_i x_i - \frac{1}{1-\mu} \left(N - \sum_{i=1}^N x_i \right) = 0$$

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

sample
mean

So MLE is data driven!

↳ requires a lot samples

MAP

$$\arg \max_{\mu} \underline{P(x|\mu)} \cdot P(\mu)$$

We need to choose/assume prior dist.
for μ , $P(\mu)$.

Consider Beta distribution

$$\text{Beta}(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot x^{\alpha-1} (1-x)^{\beta-1}$$

where $\Gamma(x) = x!$

↳ For \neq values of α and β , we
are going to have stick
prior on μ .

$$P(\mu | E) = \frac{P(E|\mu) \cdot P(\mu)}{P(E)}$$

↓

Bayes' rule

$P(E)$

is constant

posterior

$$\arg \max_{\mu} P(\mu | E) \propto \arg \max_{\mu} P(E|\mu) \cdot P(\mu)$$

$$\arg \max_{\mu} P(E|\mu) \cdot P(\mu) = \left(\prod_{i=1}^N P(x_i|\mu) \right) \cdot \frac{\Gamma(\alpha+b)}{\Gamma(\alpha) \cdot \Gamma(b)} \cdot \mu^{\alpha-1} \cdot (1-\mu)^{b-1}$$

$$\alpha \left[\prod_{i=1}^N \mu^{x_i} (1-\mu)^{1-x_i} \right] \cdot \mu^{\alpha-1} \cdot (1-\mu)^{b-1}$$

$x_i \in \{0, 1\}$

$m = \# \text{heads in event } E$

$L = \# \text{tails } " " "$

$N = m + L \quad \text{total } \# \text{coin flips}$

$$= \mu^m \cdot (1-\mu)^L \cdot \mu^{\alpha-1} \cdot (1-\mu)^{b-1}$$

$$P(\mu|E) \propto \mu^{m+\alpha-1} \cdot (1-\mu)^{L+b-1}$$

↳ Posterior has the same shape as the prior \Rightarrow Conjugate prior relationship

• What is μ using MAP?

$$\arg \max_{\mu} P(\mu | E)$$

$$\Leftrightarrow \arg \max_{\mu} \underbrace{\ln(P(\mu | E))}_{L}$$

$$L = (m + \alpha - 1) \cdot \ln(\mu) + (L + \beta - 1) \cdot \ln(1 - \mu)$$

$$\frac{\partial L}{\partial \mu} = 0 \Leftrightarrow \mu = \frac{m + \cancel{\alpha} - 1}{M + L + \cancel{\alpha} + \cancel{\beta} - 2}$$