MAP

$$\arg\min_{w} \frac{1}{2}\sum_{i=1}^{N}(t_i-y_i)^2 + \underbrace{\lambda \cdot \sum_{j=0}^{M} w_j^2}_{\text{Ridge Regularization}}$$

$$\Leftrightarrow \boxed{\arg\max_{w} \underbrace{P(x|w)}_{\text{takes a Gaussian}} \cdot \underbrace{P(w)}_{\text{takes a Gaussian shape}}}$$

Bayesian Interpretation

- we can have any distribution shape for P(x|w) and P(w).

• When we use Lasso regularizer:

the prior P(w) is Laplacian distribution

Example: Expirement flipping a coin 3 times and observing the results $\underline{H}$ $\underline{H}$ $\underline{H}$.

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

$\mu \equiv$ prob. of heads

$H \equiv 1$, $T \equiv 0$, $S = \{0, 1\}$

$$P(x = 1 \mid \mu) = \mu$$
$$P(x = 0 \mid \mu) = 1 - \mu$$

Data Likelihood:

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

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

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

$$\boxed{P(E|\mu)} = P(x_1 \cap x_2 \cap \dots \cap x_N|\mu)$$

$$\overset{\text{indep.}}{=} P(x_1|\mu) \cdot P(x_2|\mu) \cdot \dots \cdot P(x_n|\mu)$$

$$= \prod_{i=1}^{N} P(x_i|\mu)$$

$$\overset{\text{identically dist.}}{=} \boxed{\prod_{i=1}^{N} \mu^{x_i} \cdot (1-\mu)^{1-x_i}}$$

Bernoulli data likelihood

MLE

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

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

ln is a monotonic fct.

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

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

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

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

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

sample mean

So MLE is <u>data driven!</u>

↳ 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

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

where $\Gamma(x) = x!$

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

Bayes' rule

$$\underbrace{P(\mu|E)}_{\text{posterior}} = \frac{P(E|\mu).P(\mu)}{\underbrace{P(E)}_{\text{is constant}}}$$

$$\underset{\mu}{\arg\max}\, P(\mu|E) \propto \underset{\mu}{\arg\max}\, P(E|\mu).P(\mu)$$

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

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

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

$m \equiv$ # heads in event E

$L \equiv$ # tails " " "

$N = m + L$ total # coin flips

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

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

↳ Posterior has the same shape as the data likelihood ⟹ Conjugate PRIOR Relationship

- What is μ using MAP?

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

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

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

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