Bayesian learning

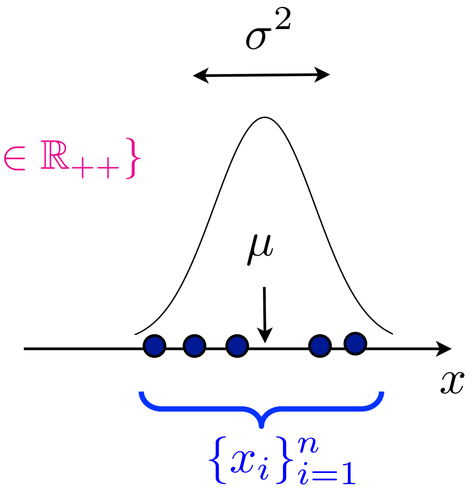
What is Bayesian learning?

Observed Data: $\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}$

Parameter: $\boldsymbol{w} = \{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{++}\}$

Model distribution: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w}) = \prod_{i=1}^{n} p(\boldsymbol{x}^{(i)}|\boldsymbol{w})$

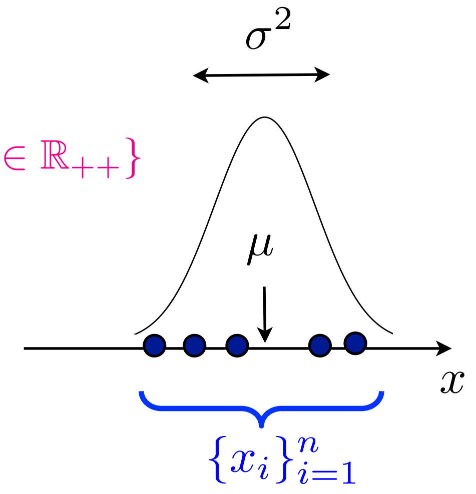
Prior distribution: $p(\boldsymbol{w}) \propto 1$



What is Bayesian learning?

Observed Data: $\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}$

Parameter: $\boldsymbol{w} = \{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{++}\}$



Model distribution: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w}) = \prod_{i=1}^{n} p(\boldsymbol{x}^{(i)}|\boldsymbol{w})$

Prior distribution: $p(\boldsymbol{w}) \propto 1$

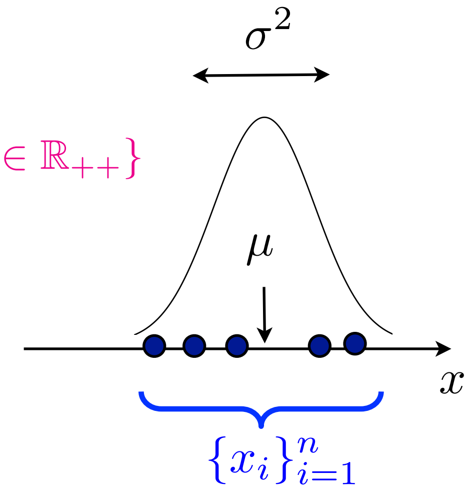
Joint distribution: $p(\underbrace{\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}}_{\text{observed}}, \underbrace{\boldsymbol{w}}_{\text{to be estimated!}}) = p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})p(\boldsymbol{w})$

We want to estimate parameter from observed data.

What is Bayesian learning?

Observed Data: $\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}$

Parameter: $\boldsymbol{w} = \{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{++}\}$



Model distribution: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w}) = \prod_{i=1}^{n} p(\boldsymbol{x}^{(i)}|\boldsymbol{w})$

Prior distribution: $p(\boldsymbol{w}) \propto 1$

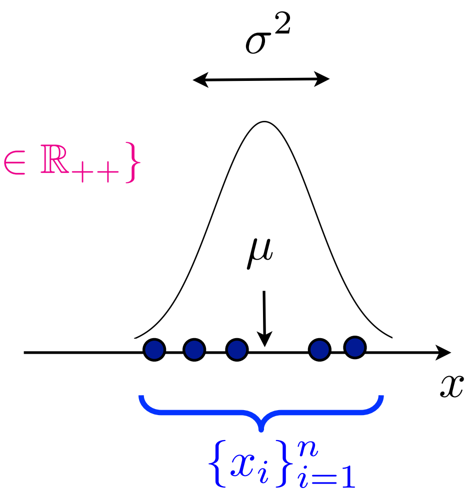
Joint distribution: $p(\underbrace{\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}}_{\text{observed}}, \underbrace{\boldsymbol{w}}_{\text{to be estimated!}}) = p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})p(\boldsymbol{w})$

$$p(b|a)p(a) = p(a,b) = p(a|b)p(b)$$

$a \to \{\boldsymbol{x}^{(i)}\}_{i=1}^{n},$

$b \to \boldsymbol{w}$

What is Bayesian learning?



Observed Data: $\{\boldsymbol{x}^{(i)}\}_{i=1}^n$

Parameter: $\boldsymbol{w} = \{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{++}\}$

Model distribution: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w}) = \prod_{i=1}^n p(\boldsymbol{x}^{(i)}|\boldsymbol{w})$

Prior distribution: $p(\boldsymbol{w}) \propto 1$

Joint distribution: $p(\underbrace{\{\boldsymbol{x}^{(i)}\}_{i=1}^n}_{\text{observed}}, \underbrace{\boldsymbol{w}}_{\text{to be estimated!}}) = p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})$

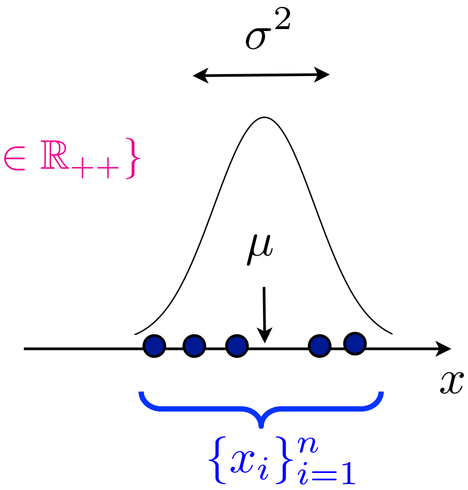
$$p(b|a)p(a) = p(a,b) = p(a|b)p(b)$$

$a \rightarrow \{\boldsymbol{x}^{(i)}\}_{i=1}^n,$

$b \rightarrow \boldsymbol{w}$

$$p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n, \boldsymbol{w}) = p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})$$

What is Bayesian learning?



Observed Data: $\{\boldsymbol{x}^{(i)}\}_{i=1}^n$ Parameter: $\boldsymbol{w} = \{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{++}\}$

Model distribution: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w}) = \prod_{i=1}^n p(\boldsymbol{x}^{(i)}|\boldsymbol{w})$

Prior distribution: $p(\boldsymbol{w}) \propto 1$

Joint distribution: $p(\underbrace{\{\boldsymbol{x}^{(i)}\}_{i=1}^n}_{\text{observed}}, \underbrace{\boldsymbol{w}}_{\text{to be estimated!}}) = p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})$

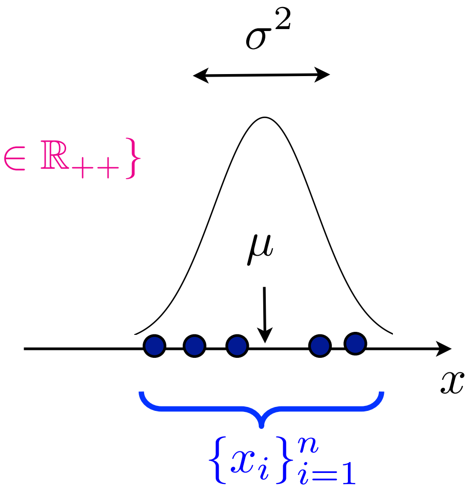
$$p(b|a)p(a) = p(a,b) = p(a|b)p(b)$$

$a \rightarrow \{\boldsymbol{x}^{(i)}\}_{i=1}^n,$

$b \rightarrow \boldsymbol{w}$

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

What is Bayesian learning?



Observed Data: $\{\boldsymbol{x}^{(i)}\}_{i=1}^n$

Parameter: $\boldsymbol{w} = \{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{++}\}$

Model distribution: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w}) = \prod_{i=1}^n p(\boldsymbol{x}^{(i)}|\boldsymbol{w})$

Prior distribution: $p(\boldsymbol{w}) \propto 1$

Joint distribution: $p(\underbrace{\{\boldsymbol{x}^{(i)}\}_{i=1}^n}_{\text{observed}}, \underbrace{\boldsymbol{w}}_{\text{to be estimated!}}) = p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})$

$$p(b|a)p(a) = p(a,b) = p(a|b)p(b)$$

$a \to \{\boldsymbol{x}^{(i)}\}_{i=1}^n,$

$b \to \boldsymbol{w}$

$$\underbrace{\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{marginal}}}_{\text{must compute}} = \underbrace{\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}}_{\text{given}}$$

Bayes posterior distribution

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

Posterior distribution:

marginal likelihood:

Bayes posterior distribution

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

Posterior distribution: $p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \dfrac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})}{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$

marginal likelihood:

Bayes posterior distribution

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

Posterior distribution: $p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \dfrac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})p(\boldsymbol{w})}{\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}_{\text{constant wrt w!}}}$

marginal likelihood:

Bayes posterior distribution

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

Posterior distribution: $p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \dfrac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})}{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$

marginal likelihood: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \displaystyle\int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$

Bayes posterior distribution

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

Posterior distribution: $$p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \frac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})}{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$$

marginal likelihood: $$p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$$

Bayes estimator: $$\widehat{\boldsymbol{w}} = \int \boldsymbol{w} \cdot \underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{posterior}}d\boldsymbol{w} = \langle \boldsymbol{w} \rangle_{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$$

Want to know the parameter $\boldsymbol{w}$ estimated from observation $\{\boldsymbol{x}^{(i)}\}_{i=1}^n$.

Bayes posterior distribution

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

Posterior distribution: $p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \dfrac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})p(\boldsymbol{w})}{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}$

marginal likelihood: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \displaystyle\int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$

Bayes estimator: $\widehat{\boldsymbol{w}} = \displaystyle\int \boldsymbol{w} \cdot p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})d\boldsymbol{w} = \langle \boldsymbol{w} \rangle_{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}$

Predictive distribution: $p(\boldsymbol{x}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \displaystyle\int p(\boldsymbol{x}, \boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})d\boldsymbol{w}$

Want to predict a future $\boldsymbol{x}$ estimated from past observation $\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}$.

Bayes posterior distribution

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

Posterior distribution: $p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \dfrac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})}{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$

marginal likelihood: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$

Bayes estimator: $\widehat{\boldsymbol{w}} = \int \boldsymbol{w} \cdot p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)d\boldsymbol{w} = \langle \boldsymbol{w} \rangle_{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$

Predictive distribution: $p(\boldsymbol{x}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \int p(\boldsymbol{x}, \boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)d\boldsymbol{w}$

$$\Big\downarrow \quad p(a,b) = p(a|b)p(b)$$

$$= \int p(\boldsymbol{x}|\boldsymbol{w}, \{\boldsymbol{x}^{(i)}\}_{i=1}^n)p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)d\boldsymbol{w}$$

Bayes posterior distribution

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

Posterior distribution: $p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \dfrac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})p(\boldsymbol{w})}{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}$

marginal likelihood: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$

Bayes estimator: $\widehat{\boldsymbol{w}} = \int \boldsymbol{w} \cdot p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})d\boldsymbol{w} = \langle \boldsymbol{w} \rangle_{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}$

Predictive distribution: $p(\boldsymbol{x}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \int p(\boldsymbol{x}, \boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})d\boldsymbol{w}$

independent!

$$= \int p(\boldsymbol{x}|\boldsymbol{w}, \{\boldsymbol{x}^{(i)}\}_{i=1}^{n})p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})d\boldsymbol{w}$$

Bayes posterior distribution

$$\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{posterior}}\underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{marginal}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n, \boldsymbol{w})}_{\text{joint}} = \underbrace{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})}_{\text{likelihood}}\underbrace{p(\boldsymbol{w})}_{\text{prior}}$$

Posterior distribution: $$p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \frac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})}{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$$

marginal likelihood: $$p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^n|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$$

Bayes estimator: $$\widehat{\boldsymbol{w}} = \int \boldsymbol{w} \cdot p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)d\boldsymbol{w} = \langle \boldsymbol{w} \rangle_{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}$$

Predictive distribution: $$p(\boldsymbol{x}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n) = \int p(\boldsymbol{x}, \boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)d\boldsymbol{w}$$

independent!

$$= \int p(\boldsymbol{x}|\boldsymbol{w}, \{\boldsymbol{x}^{(i)}\}_{i=1}^n)p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)d\boldsymbol{w}$$

$$= \int \underbrace{p(\boldsymbol{x}|\boldsymbol{w})}_{\text{model distribution}}\underbrace{p(\boldsymbol{w}|\{\boldsymbol{x}^{(i)}\}_{i=1}^n)}_{\text{posterior}}d\boldsymbol{w}$$

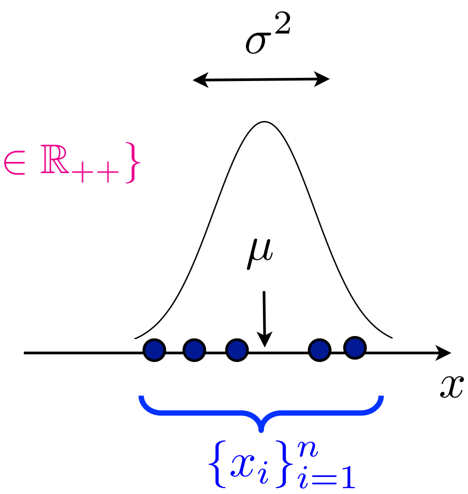
Bayesian learning

Observed Data: $\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}$

Parameter: $\boldsymbol{w} = \{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{++}\}$

Model distribution: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n} | \boldsymbol{w}) = \prod_{i=1}^{n} p(\boldsymbol{x}^{(i)} | \boldsymbol{w})$

Prior distribution: $p(\boldsymbol{w}) \propto 1$



Posterior distribution: $p(\boldsymbol{w} | \{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \dfrac{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n} | \boldsymbol{w}) p(\boldsymbol{w})}{p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}$

marginal likelihood: $p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \int p(\{\boldsymbol{x}^{(i)}\}_{i=1}^{n} | \boldsymbol{w}) p(\boldsymbol{w}) d\boldsymbol{w}$

Bayes estimator: $\widehat{\boldsymbol{w}} = \int \boldsymbol{w} \cdot p(\boldsymbol{w} | \{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) d\boldsymbol{w} = \langle \boldsymbol{w} \rangle_{p(\boldsymbol{w} | \{\boldsymbol{x}^{(i)}\}_{i=1}^{n})}$

Predictive distribution: $p(\boldsymbol{x} | \{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) = \int p(\boldsymbol{x} | \boldsymbol{w}) p(\boldsymbol{w} | \{\boldsymbol{x}^{(i)}\}_{i=1}^{n}) d\boldsymbol{w}$

Questions?