

[Bayesian Inference]

α : a data point
 θ : the parameter of the data point's distribution $x \sim P(x|\theta)$
 α : the hyperparameter of the parameter $\theta \sim P(\theta|\alpha)$
 X : a set of n observed data points $x_1, x_2, x_3, \dots, x_n$
 \tilde{x} : a new data point whose distribution is to be predicted.

ML (maximum likelihood)

$$\hat{\theta} = \arg\max_{\theta} f(x|\theta)$$

MAP (maximum a posteriori estimation)

$$\hat{\theta} = \arg\max_{\theta} f(\theta|x)$$

likelihood
sampling distribution

$$L(\theta|X) = P(X|\theta)$$

$$P(X|\theta) = \prod_k P(x_k|\theta)$$

prior distribution $P(\theta|\alpha)$
prior

$$P(\theta|x) =$$

$$\frac{P(X|\theta) \cdot P(\theta|\alpha)}{P(X)}$$

posterior

posterior distribution

$$P(\theta|x, \alpha) = \frac{P(X|\theta) \cdot P(\theta|\alpha)}{P(X|\alpha)}$$

model evidence

marginal likelihood

$$\begin{aligned}
 P(X|\alpha) &= \sum_{\theta} P(X|\theta) \cdot P(\theta|\alpha) \\
 &= \int_{\theta} P(X|\theta) \cdot P(\theta|\alpha) d\theta \\
 &= \underline{\underline{f(\theta)}}
 \end{aligned}$$

[Bayesian Prediction]

posterior predictive distribution

$$P(\tilde{x}|X, \alpha) = \int_{\theta} P(\tilde{x}|\theta) \cdot P(\theta|x, \alpha) d\theta$$

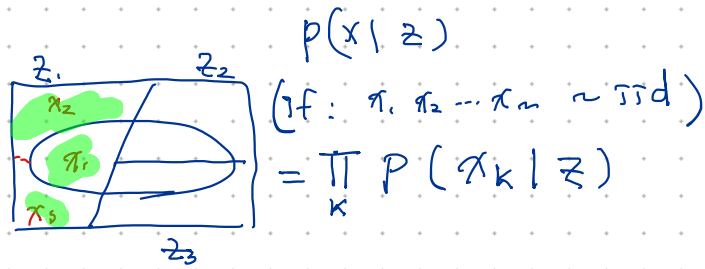
prior predictive distribution

$$P(\tilde{x}|\alpha) = \int_{\theta} P(\tilde{x}|\theta) \cdot P(\theta|\alpha) d\theta$$

predictive inference

predict the distribution of new, unobserved data point

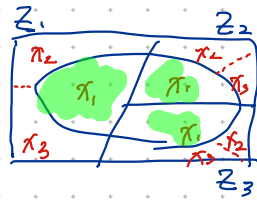
[Naïve Bayes]



z
 \downarrow

x

$$p(z|x) = \frac{p(x|z) \cdot p(z)}{p(x)}$$



$$= \sum_z p(x, z)$$

$$= \sum_z p(x|z) p(z)$$

$$= \int p(x|z) p(z) dz$$