[Bayesian Inference]

α

θ

X x $\tilde{x}$

- x : 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 $x_1\ x_2\ x_3 \cdots x_n$
- $\tilde{x}$: a new data point whose distribution is to be predicted.

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

MAP (maximum posterior estimation)
$\hat{\theta} = \arg\max_{\theta} f(\theta|x)$

likelihood $L(\theta|X) = p(X|\theta)$
sampling distribution

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

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

$$p(\theta|X) = \frac{p(X|\theta)\cdot p(\theta)}{p(X)}$$

posterior
posterior distribution

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

model evidence
marginal likelihood

$$p(X|\alpha) = \sum_i p(X|\theta_i)\, p(\theta_i|\alpha)$$
$$= \int_\theta p(X|\theta)\cdot p(\theta|\alpha)\, d\theta$$

$\underline{\underline{f(\theta)}}$

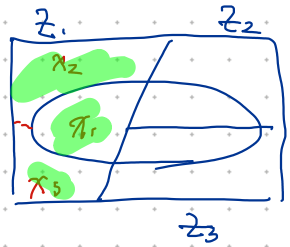
[Bayesian Prediction]

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

$p(x \mid z)$



(if: $x_1\ x_2 \cdots x_m \sim$ iid)

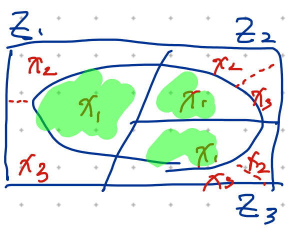
$= \prod_k p(x_k \mid z)$ Naive Bayes

z

$\downarrow$

x

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



$= \sum_i p(x, z_i)$

$= \sum_i p(x \mid z_i)\, p(z_i)$

$= \int p(x \mid z)\, p(z)\, dz$