THE UNIVERSITY OF CHICAGO
UChicago Medicine | Biological Sciences Division

ZeD

CCTS 40500: Machine Learning & Advanced Analytics For Biomedicine

Bayesian Statistics: Posterior ∝ Likelihood × Prior

Ishanu Chattopadhyay (ishanu@uchicago.edu)

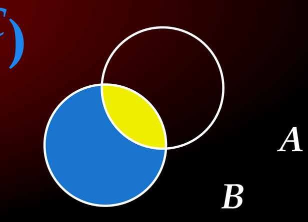
$$Pr(H|E) = \frac{Pr(E|H)Pr(H)}{\sum_{H \in \mathbb{H}} Pr(E|H)Pr(E)} = \frac{Pr(E|H)Pr(H)}{Pr(E)}$$

Posterior Likelihood Prior

$$Pr(H|E) = \frac{Pr(E|H)Pr(H)}{Pr(E|H)Pr(E) + Pr(E|\neg H)Pr(\neg H)}$$

Marginal Likelihood

$$Pr(A|B) = \frac{Pr(B \cap A)}{Pr(B)} = \frac{Pr(B \cap A)}{Pr(B \cap A) + Pr(B \cap A^C)}$$



Concepts:

- Bayesian Estimation/Inference
- Bayesian Prediction
- Bayes Classifier
- Bayes Estimator
- Bayes Risk/Error
- Conjugate Priors
- MLE and MAP

Bayesian Nomenclature:

$x \sim P(x\|\theta)$	Prior distribution, – θ is a parameter
$\theta \sim P(\theta\|\alpha)$	Parameters are random varaibles, – α is a hyper – parameter
$P(x\|\theta)$	Sampling distribution, same as Likelihood
$L(\theta\|x)$	Likelihood (same as sampling distribution)
$P(x\|\alpha)$	Marginal Likelihood
$P(\theta\|x,\alpha)$	Posterior

Equations (with hyperparameter explicit):

$$P(x|\alpha) = \int P(x|\theta)P(\theta|\alpha)d\theta$$

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

And, $\hat{x}$ is the prediction:

$$P(\hat{x}|x,\alpha) = \int P(\hat{x}|\theta)P(\theta|x,\alpha)d\theta$$

$$P(\hat{x}|\alpha) = \int P(\hat{x}|\theta)P(\theta|\alpha)d\theta$$

Classification (Feature vector: $x \in \mathbb{X}$, labels: $y \in \mathbb{Y}$)

$P(x,y)$	probability feature vector x has label y
$h : \mathbb{X} \to \mathbb{Y}$	Classification Rule
$R(h) = P(h(x) \neq y)$	Risk or probability of error

Bayes Classifier:

$$h^\star(x) = \arg\max_{y \in \mathbb{Y}} P(y|x)$$

Theorem (Optimality of Bayes Classifier):
For any classifiction rule h,

$$R(h) \geqq R(h^\star)$$

Bayes Estimator & Optimal Risk

$\theta \in \Theta$	To be Estimated
$X \in \mathbb{X}$	Observed data
$L(\theta,\hat{\theta})$	Loss Function
$h : \mathbb{X} \to \Theta$	Estimator
$R(L,h) \triangleq \mathbf{E}[L(\theta,h(X))]$	Average Risk

Objective: Minimize Average Risk (Note, it depends on the loss function)

Bayes Estimator:

$$h^\star(X) = \arg\inf_h R(L,h)$$

Bayes Risk: $R(L,h^\star)$

Bayes Estimator For Different Loss Functions

$$L(\theta,\hat{\theta}) = |\theta - \hat{\theta}| \Rightarrow h^\star(X) = \text{median } Pr(\theta|X)$$

$$L(\theta,\hat{\theta}) = (\theta - \hat{\theta})^2 \Rightarrow h^\star(X) = \mathbf{E}[Pr(\theta|X)]$$

$$L(\theta,\hat{\theta}) = \omega(\theta - \hat{\theta})^2 \Rightarrow h^\star(X) = \frac{\mathbf{E}[Pr(\omega\theta|X)]}{\mathbf{E}[Pr(\omega|X)]}$$

$$L(\theta,\hat{\theta}) = \delta_{\theta,\hat{\theta}} \Rightarrow h^\star(X) = \arg\max_\theta Pr(\theta|X)$$

Nearest Neighbor Classifier

Decision Rule:

$$h(x) = \text{mode}(\{y'' : (x'',y'') \in S_x\})$$

where S_x is defined as

$$d(x,x') \geqq \max_{(x'',y'') \in S_x} d(x,x'')$$

$$R\left(h^\star\right) \leqq R\left(h^{NN}\right) \leqq R\left(h^\star\right)\left(2 - \frac{c}{c-1}R(h^\star)\right)$$

Naive Bayes Classifier

Objective: Estimate $Pr(x|y) = \prod_k Pr(x_k|y)$ with the assumption that the features x_k are independent

$$h(X) = \log Pr(y) + \arg\max_y \sum_k \log Pr(x_k|y)$$

Logistic Regression

Objective: Estimate $Pr(y_k|x)$ assuming:

$$Pr(y_k|x) = \frac{1}{1 + e^{-y(\omega^T x + b)}}$$