Bayes Classifier

July 7, 2023

Bayes Rule

- A and B are two events, then

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Counting in two ways

- $P(A), P(B)$ - Proportion of times event A, B occured

Counting in two ways

- $P(A), P(B)$ - Proportion of times event A, B occured
- $P(B|A)$ - Given A occured, proportion of times B occured

Counting in two ways

- $P(A), P(B)$ - Proportion of times event A, B occured
- $P(B|A)$ - Given A occured, proportion of times B occured
- $P(A|B)$ - Given B occured, proportion of times A occured

Counting in two ways

- $P(A), P(B)$ - Proportion of times event A, B occured
- $P(B|A)$ - Given A occured, proportion of times B occured
- $P(A|B)$ - Given B occured, proportion of times A occured
- $P(A)P(B|A)$ - Proportion of times both A and B occured

Counting in two ways

- $P(A), P(B)$ - Proportion of times event A, B occured
- $P(B|A)$ - Given A occured, proportion of times B occured
- $P(A|B)$ - Given B occured, proportion of times A occured
- $P(A)P(B|A)$ - Proportion of times both A and B occured
- $P(B)P(A|B)$ - Proportion of times both A and B occured

Counting in two ways

- $P(A), P(B)$ - Proportion of times event A, B occured
- $P(B|A)$ - Given A occured, proportion of times B occured
- $P(A|B)$ - Given B occured, proportion of times A occured
- $P(A)P(B|A)$ - Proportion of times both A and B occured
- $P(B)P(A|B)$ - Proportion of times both A and B occured

$$\implies P(B|A)P(A) = P(A|B)P(B)$$

Bayesian Classification

Loan Lending Problem

Figure: Could you lend me a loan?

Classification/Prediction Sub Problem

- Predict when a customer asks for a loan whether he will default or return? Minimize error

Classification/Prediction Sub Problem

- Predict when a customer asks for a loan whether he will default or return? Minimize error
- What is predicton/misclassification error?

Classification/Prediction Sub Problem

- Predict when a customer asks for a loan whether he will default or return? Minimize error

- What is predicton/misclassification error?

customer returns when we predict he will default
or customer defaults when we predict he will return

Classification/Prediction Sub Problem

- Predict when a customer asks for a loan whether he will default or return? Minimize error
- What is predicton/misclassification error?

customer returns when we predict he will default
or customer defaults when we predict he will return

- Known information

P(Default)	0.2
P(Return)	0.8

Table: Prior

Prediction Strategies

- Predict the customer will always ‘Default’

Prediction Strategies

- Predict the customer will always ‘Default’
- Predict the customer will always ‘Return’

Prediction Strategies

- Predict the customer will always 'Default'
- Predict the customer will always 'Return'
- Toss your lucky coin with bias for head p if head - predict he defaults and tail - predict he returns

Prediction Error in Strategies

- 'Say always default' : $P(\text{error}) = P(\text{whenever return happens}) = 0.8$

Prediction Error in Strategies

- 'Say always default' : $P(\text{error}) = P(\text{whenever return happens}) = 0.8$
- 'Say always return' : $P(\text{error}) = P(\text{whenever default happens}) = 0.2$

Prediction Error in Strategies

- 'Say always default' : $P(\text{error}) = P(\text{whenever return happens}) = 0.8$
- 'Say always return' : $P(\text{error}) = P(\text{whenever default happens}) = 0.2$
- Mixed Strategy - $P(\text{error}) = p * 0.8 + (1 - p) * 0.2 = 0.2 + 0.6p$

How to incorporate information?

- Suppose we have additional information

How to incorporate information?

- Suppose we have additional information
- We call a loan High risk H - if Loan amount $>$ 50% annual salary
- We call a loan Low risk L - if Loan amount $<$ 50% annual salary
- Additional Information

$P(H\|D)$	2/3
$P(H\|R)$	1/10

Table: Likelihood

Posterior Probability Computation

- $P(D|H)$ (posterior) $= \frac{P(H|D) \text{ (Likelihood) } * P(D) \text{ (Prior)}}{P(H) \text{ (Evidence)}}$

Posterior Probability Computation

- $P(D|H)$ (posterior) $= \frac{P(H|D) \text{ (Likelihood) } \quad * P(D) \text{ (Prior)}}{P(H) \text{ (Evidence)}}$

$$P(H|D)P(D) = 2/3 * 2/10 = 4/30$$

$$\begin{aligned} P(H) &= P(H|D)P(D) + P(H|R)P(R) \\ &= 2/3 * 2/10 + 1/10 * 8/10 = 64/300 \end{aligned}$$

$$P(D|H) = 4/30 * 300/64 = 5/8 = 0.625$$

- Similarly we can compute $P(D|L) = 0.08$

Bayes Classification Rule

- How to predict with new information?

Bayes Classification Rule

- How to predict with new information?

- Since $P(D|H) > P(R|H)$, predict default for high risk loans

Bayes Classification Rule

- How to predict with new information?
- Since $P(D|H) > P(R|H)$, predict default for high risk loans
- Since $P(R|L) > P(D|L)$, predict return for low risk loans

Bayes Classification Rule

- How to predict with new information?
- Since $P(D|H) > P(R|H)$, predict default for high risk loans
- Since $P(R|L) > P(D|L)$, predict return for low risk loans
- What is misclassification error?

Bayes Classification Rule

- How to predict with new information?
- Since $P(D|H) > P(R|H)$, predict default for high risk loans
- Since $P(R|L) > P(D|L)$, predict return for low risk loans
- What is misclassification error?

$$\begin{aligned} P(\text{error}) &= P(\text{person defaults and we predict return}) \\ &+ P(\text{person returns and we predict default}) \\ &= P(D)P(L|D) + P(R)P(H|R) = 0.15 \end{aligned}$$

Bayes Classification Rule

- How to predict with new information?
- Since $P(D|H) > P(R|H)$, predict default for high risk loans
- Since $P(R|L) > P(D|L)$, predict return for low risk loans
- What is misclassification error?

$$\begin{aligned} P(\text{error}) &= P(\text{person defaults and we predict return}) \\ &\quad + P(\text{person returns and we predict default}) \\ &= P(D)P(L|D) + P(R)P(H|R) = 0.15 \end{aligned}$$

- Notice $P(\text{error}) = 0.15 < 0.2$ as we used more information.

Bayes Classification Summary

- In general, there are more than one feature

Bayes Classification Summary

- In general, there are more than one feature
- Likelihood information is given as density/distribution over features under a given class

Bayes Classification Summary

- In general, there are more than one feature
- Likelihood information is given as density/distribution over features under a given class
- Bayes rule predicts the class which has high posterior probability given that feature

Bayes Classification Summary

- In general, there are more than one feature
- Likelihood information is given as density/distribution over features under a given class
- Bayes rule predicts the class which has high posterior probability given that feature
- Bayes rule is optimal and achieves the least classification error

Questions

Thank you !

References I

Roberto Calandra, André Seyfarth, Jan Peters, and Marc Peter Deisenroth. 2016. Bayesian optimization for learning gaits under uncertainty. *Annals of Mathematics and Artificial Intelligence* 76, 1-2 (2016), 5–23.

Marc Deisenroth and Carl E Rasmussen. 2011. PILCO: A model-based and data-efficient approach to policy search. In *Proceedings of the 28th International Conference on machine learning (ICML-11)*. 465–472.

Daniel J Lizotte, Tao Wang, Michael H Bowling, and Dale Schuurmans. 2007. Automatic Gait Optimization with Gaussian Process Regression.. In *IJCAI*, Vol. 7. 944–949.

Ruben Martinez-Cantin. 2017. Bayesian optimization with adaptive kernels for robot control. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 3350–3356.

Ruben Martinez-Cantin, Nando de Freitas, Eric Brochu, José Castellanos, and Arnaud Doucet. 2009. A Bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot. *Autonomous Robots* 27, 2 (2009), 93–103.

References II

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. 2012. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*. 2951–2959.

Aaron Wilson, Alan Fern, and Prasad Tadepalli. 2014. Using trajectory data to improve Bayesian optimization for reinforcement learning. *The Journal of Machine Learning Research* 15, 1 (2014), 253–282.