Oct 13th

# Naive Bayes

BAYES RULE $P(C|A) = \frac{P(A|C)\,P(C)}{P(A)}$ (P(C): prior; P(A): prior)

BAYESIAN CLASSIFIER

given record w/ attributes $A_1, A_2 \cdots A_n$

goal: predict class C → value of C that maximizes $P(C \mid A_1, A_2 \cdots A_n)$

$$P(C \mid A_1, A_2 \cdots A_n) = \frac{P(A_1, A_2 \cdots A_n \mid C)\,P(C)}{P(A_1, A_2 \cdots A_n)}$$

*NAIVE BAYES SAYS WE SHOULD MAXIMIZE NUMERATOR.

↳ so find records in class C and assuming feature independence (NOT ALWAYS CASE) find $Pr(A_1, A_2 \cdots A_n \mid C)$.

↳ THEN WE CAN SAY $P(A_1|C)\,P(A_2|C) \cdots P(A_n|C)$

- DISCRETE
  - $P(C) = N_C/N$
  - $P(A_i|C_k) = |A_i|_{C_k} / N_{C_k}$ where $|A_i|_{C_k}$ → # of instances w/ attr $A_i$ and belong to class $C_k$.
- CONTINUOUS
  - Discretize → 1 ordinal attr / bin. violates II assumption
  - two way split
  - Prob density estimation.

NOTE: if we haven't seen a particular combo of records → leads to a 0 value.

↳ USE TRANSFORMATIONS

LAPLACE

M-ESTIMATE

Predict the class that maximizes the probability of being in that class given the attributes.

# Support Vector Machines

find the widest street that separates our classes → maximize split b/w 2 classes.

↳ line in the middle is the decision boundary

find a vector perpendicular to decision boundary, take dot prod and if its greater than some constant then predict class +

*DECISION RULE $\vec{w} \cdot \vec{u} + b \geq 0$ then +

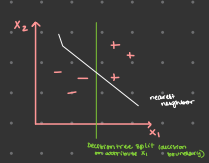
need to find $\vec{w}, b$.

NOTE: For an unknown $\vec{u}$ $-1 < \vec{w} \cdot \vec{u} + b < 1$ ~> create the widest street $\vec{w} \cdot \vec{x}_+ + b \geq 1$, $\vec{w} \cdot \vec{x}_- + b \leq -1$

$y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1$ → $y_i(\vec{w}\vec{x}_i + b) - 1 = 0$

WIDTH $= (\vec{x}_+ - \vec{x}_-) \cdot \frac{\vec{w}}{\|\vec{w}\|}$

Points for side of street fully describe the model.




When you need to change perspective apply a transformation

via kernel trick.