Naive Bayes Classifier

All previous formulations of Classification =

given data $\{x_{j1}, x_{j2}, \dots, x_{jD}\}$, $\forall d$, $d = 1, 2, \dots, D$

Predict $\hat{y}_j$: $\hat{y}_j \subset \{C_k, \forall k, k = 1, 2, \dots, K\}$

Alternative but equiv formulation:

given data ...

what is the prob of a class given inputs

$p(C_k \mid x_{j1}, x_{j2}, \dots, x_{jd})$, $\forall k$ $k = 1, 2, \dots, K$

$\hat{y}_j = C_k$ for $\operatorname{argmax}\ p(C_k \mid x_{j1}, x_{j2}, \dots, x_{jD})$

Bayes Rules:

for two events A & B

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

Bayes Rules

for two events A & B

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

$P(B|A)$ = cond prob of seeing B given observation of A

$P(A)$

$P(B)$ = Prior probability of B or

$P(A|B)$ = Prob of A occuring given obs of B

Probabalistic Model formulation

Rewrite bayes rule for a set of possible classes

$$P(y=C_k|X) = \frac{P(X|y=C_k)P(Y=C_k)}{P(X)}$$

$P(X)$ can be ignored → $P(X)$

as it is the same for ALL classes

Same as:

$$P(y=C_k|X) = P(X_1=x_{j1}, x_2=x_{j2}, \ldots, X_D=x_{jD}|y=C_k)$$
$$P(Y=C_k)$$

BAYES Classifier

$$P(C_k|X) = P(X_{j1}, X_{j2}, \ldots, X_{jD} | C_k) P(C_k)$$

$P(C_k)$ can be estimated from the training data as $\frac{N_{Ck}}{N}$ → N of C_k instan
→ n of instance

in eng: the prob of seeing class C_k given the parameters, against the overall prob of seeing the class

$P(X_{jD}|C_k)$ can be estimated FROM the training data given a distris e.g gaussian $N(\mu_{C_k}, \Sigma_{C_k})$

Probability of X given the parameter distribution

$$P(X_{jD}|C_k) = E[P(X|\mu_{C_k}, \Sigma_{C_k})]$$

μ_{Ck} = D-Dimension mean vector — mean
Σ_{Ck} = D x D covariance — SD
$E[\cdot]$ = Expected value

Find Optimal Params

Assuming a gaussian distribution $N(\mu_{ck}, \Sigma_{ck})$ the prob dist params are μ_{c_k}, Σ_{c_k}

the optimal dist params max the likl.hood that the given data $= E[P(X|\mu_{ck}, \Sigma_{ck})]$

Recall that opt value of Param θ for a function $L = \frac{\partial L}{\partial \theta} = 0$

Log likihood derivation

$P(X|\mu_{ck}, \Sigma_{ck})$ = Plug in gaussian prob Dens func

(1) take log of both sides
(2) log of Product = log of sum
(3) apply log rules

finally max/opt likilihood of X given params μ_{ck} Σ_{ck}

$$\frac{\partial}{\partial u} \log P(x|\mu_{ck}, \Sigma_{ck}) = 0$$

- Sub in log distribution $\log P(x|\mu_{ck}, \Sigma_{ck})$
- remove non-μ terms as per Diff rule
- Expand & Simplify again

- After full simplification, make μ_{ck} the subject

$$\mu_{ck} = \frac{1}{N_{ck}} \sum^{N_{ck}} x_n$$

↳ Ends up being standard mean formula

- Next optimize $\Sigma_{ck} = \frac{\partial}{\partial \Sigma} \log P(X/\mu_{ck}, \Sigma_{ck}) = 0$
- Plug in logs, expand, simplify, set to Σ

$$\Sigma_{ck} = \frac{1}{N_{ck}} \Sigma (x_n - \mu_{ck})^T (x_n - \mu_{ck})$$

↳ Ends up as standard variance formula

Now we have optimal μ & Σ values

$P(X|\mu_{ck}, \Sigma_{ck})$ = Prob Dens func plug in

$\mu_{ck} = \frac{1}{N_{ck}} \sum x_n$

$\Sigma_{ck} = \frac{1}{N_{ck}} \sum (x_n - \mu_{ck})^T (x_n - \mu_{ck})$

Plug optimal values into PDF = Bayes Class

Naive Bayers classifier assumes input params are independent so can be isolated

Summary

Bayes Classifier gives the posterior probability of each class by updating the prior probability of the class with the likelihood of the observed data