

Bayes Optimal Classifier, Naive Bayes

NB, Now Available in
Multinomial and Gaussian flavors

10-701 Recitation, CMU

9/20/2007

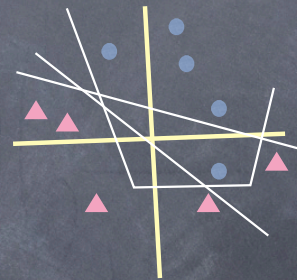
Sue Ann Hong

Classification

- Goal: Learn the ^{parameters, to} underlying function

f : features $x \rightarrow$ class y

e.g.: words \rightarrow "spam"



- many many possible functions...
"hypotheses"

- "hypothesis space"

- e.g. functions with hyperplane (\sim line)
boundaries "linear classifiers"

f: features_x → class_y

- One way to learn f:

- find $P(Y|X)$

- classifier?

- $$f(X) = \operatorname{argmax}_k P(Y=k|X)$$

- why?

- Bayes optimality

- if we know $P(Y|X)$ accurately, the classifier above is optimal.

Finding $P(Y|X)$

“Generative” learning

- Bayes Rule:

- $$P(Y|X) = P(X|Y) P(Y) / P(X)$$

- $$\propto P(X|Y) P(Y)$$

- Learn $P(X|Y)$, $P(Y)$

- Can also “generate” the data from them

- Another way: learn $P(Y|X)$ directly

- e.g. Logistic regression

- “discriminative”

Finding $P(X|Y)$, $P(Y)$

"Generative" learning, Bayes optimal classifier

- 3 classes = {spam, not spam, maybe}
- 10,000 binary features = {"CA\$H", "Rolex", ...}
- # params for $P(Y)$?
 - $P(Y=\text{spam}), P(Y=\text{not spam}) \rightarrow 2$
- # params for $P(X|Y) = P(X_1, \dots, X_n|Y)$?
 - $P(\text{"CA$H"}=1, \text{"Rolex"}=1, \dots | Y=\text{spam})$
 - $P(\text{"CA$H"}=1, \text{"Rolex"}=1, \dots | Y=\text{not spam})$
 - $P(\text{"CA$H"}=1, \text{"Rolex"}=1, \dots | Y=\text{maybe})$
 - $x (2^{10,000}-1) \sim 6^{10,000} = \dots$

$$6^{10,000} = 10^{7781.51} =$$

```
325464658549366205883184429132500573463884637762028178443836042858527503193176\
637401787736300238202768004806922154307358113527295872833268827614109076375493\
419874332840764232553207490334347503015832945623789498616297577964397267434720\
956828790319673771248105387488832042240721213353459735500782713467052102165658\
343759175809709868584178546308480596843608568545571399805365001369520476147372\
707586211182859659854674752704195940975223903004928036333858149683386177728165\
964220863893599109666571725039391538649256624517768370928780248122994233332737\
918001208535649132595979125488271969760907793552781682875739025924894254138262\
911829255875367614193406645317610033759099796497068127897034255583684455430912\
03642304915633452211159655481083956888093017116443171228567594129974715691991\
487152272451836859907936071851942824088633028530649564379674012613417221809574\
294897719514701566102297521240068189093467946479598047715158941212594582001098\
780096433962861203337358324891053224168618860958558406805404230965379691080135\
562164961632469165075821725836330108835561479783120364165214266521428999432929\
907648843895195863965605887943327940678157648642235132172271544812090360945205\
281955609398307472632187789258160421902536044672398179280112191451984770683814\
926810159703553940777814278087903754090638663565299503771608082493890361058099\
898377917330016931082495286917359637953305980763970420653701965229329506589850\
103007849446028066322998658336514271246470669855444247951369926252837200593369\
885322462130469115746480057513877669740922144533747296554283134416433907274093\
670940540767645563312693693321475420942189246449372017441250111206912854336996\
2603724221352258720874288820605708325364679811754764930000316976345863970863719\
687990272517988426846696710153277664543141610026187867233599555468053287281962\
32723803645155993635425877328998423673043758851339859032227768161445621141298\
087370599350128496857013368967806341414569752272132315582129145303354965536279\
98948615514868157964022438000218271151966983687794231423035788917989915040685\
195978034947032257202816037658815106637855820868184702675838319820661750229967\
980362014165212119358708963986993755144041152091139449862493878892445946741594\
54554223815458848805927132395259678060986003165632255138021877191064115459271\
825130390240958747664958766965034079021122102107529255421657038560947463018167\
354876174227528507245225883170615839283452018740776886738288723555497250041263\
133761154707592420552617697589171773852967182671116858148562028830287101101763\
571007901653259040738340828832725663689161683424213814771012913690892776181086\
923240108753179378009228232687298506301750558671391385394155718382663347275115\
923295553406609219845554445034251472597164738004948276007101485141652029540725\
229833261340262001541730403418332116247563571556452768940448116351607431273748\
57336674812152342095977638177388208737930119317106591385874241832968062872396\
087587233425257074675340442177101375103016272862039022026984228962024883890396\
927036103167808126874962162118557945800223947855677081430205817090871764577476\
749083729718229030415710011517856275218619737163415572785712887767066308139375\
```

Finding $P(X|Y)$, $P(Y)$

"Generative" learning: Naive Bayes

- Introducing, conditional independence

$$P(X_1, X_2 | Y) = P(X_1 | Y) P(X_2 | Y)$$

- Given Y , X_1 tells me nothing about X_2

$$\begin{aligned} P(Y|X) &= P(X|Y) P(Y) / P(X) \\ &= P(X_1|Y) \dots P(X_n|Y) P(Y) / P(X) \\ &= \prod_i P(X_i|Y) P(Y) / P(X) \end{aligned}$$

- Learn $P(X_1|Y)$, ..., $P(X_n|Y)$, $P(Y)$

Finding $P(X|Y)$, $P(Y)$

"Generative" learning: Naive Bayes

- 3 classes = {spam, not spam, maybe}
 - 10,000 binary features = {"bored", "CA\$H", ...}
 - # params for $P(Y)$
 - $P(Y=\text{spam})$, $P(Y=\text{not spam}) \rightarrow 2$ vs. 2
 - # params for $P(X|Y) = P(X_1|Y) \dots P(X_n|Y)$
 - $P(\text{"CA$H"}=1 | Y=\text{spam})$
 - $P(\text{"CA$H"}=1 | Y=\text{not_spam})$
 - $P(\text{"CA$H"}=1 | Y=\text{maybe})$
- $\times 10,000 = 30,000 \sim 10^4$ vs. $10^{7781.51}$

Naive Bayes

- why?
 - # params
 - "simpler" - less likely to overfit!
- Applet time!
<http://www.cs.technion.ac.il/~rani/LocBoost/>

= Intermission =

- So far, we've been treating $P(X_i=j | Y=k)$, $P(Y=k)$ as one parameter each.
 - True for the counting Naive Bayes from class
 - But not necessarily
 - Because we can assign any reasonable function form to $P(X|Y)$ and $P(Y)$ for generative learning

Finding $P(X_i|Y)$, $P(Y)$

Two Common Flavors for $P(X_i|Y)$

• $P(Y) \sim \text{Multinomial}(\theta_1, \dots, \theta_k)$

• $P(Y=c) = \theta_c$

MLE for θ : counting

• Multinomial $P(X_i|Y)$

$P(X_i|Y=c) \sim \text{Multinomial}(\tau_{ci1}, \tau_{ci2}, \dots)$

$P(X_i=v|Y=c) = \tau_{civ}$

MLE for τ : counting

• Gaussian $P(X_i|Y)$

$P(X_i|Y=c) \sim N(\mu_{ci}, \sigma_{ci}^2)$

$P(X_i=v|Y=c) = \frac{1}{\sqrt{2\pi\sigma_{ci}}} \exp\left[-\frac{1}{2\sigma_{ci}^2}(v - \mu_{ci})^2\right]$

MLE for μ_{ci}, σ_{ci}^2
= sample mean & var

classifying <><

See Mathematica notebook

Smelly	Color	Temperature	Goodness
yes	green	cool	bad
no	white	cold	good
yes	red	cool	sellable
yes	red	cold	good
yes	white	cool	good
yes	red	hot	bad

f1

yes	red	cool	
-----	-----	------	--

f2

yes	white	cool	
-----	-------	------	--

XOR

X1	X2	Y
1	0	1
0	1	1
1	1	0
0	0	0

conditional independence?

Naive Bayes

$$P(Y=1) = ?$$

$$P(Y=0) = ?$$

$$P(X1=1|Y=0) = ?$$

$$P(X2=1|Y=0) = ?$$

$$P(X1=1|Y=1) = ?$$

$$P(X2=1|Y=1) = ?$$

$$P(X1=1, X2=0|Y=0) = ?$$

$$P(X1=0, X2=1|Y=0) = ?$$

$$P(X1=0, X2=0|Y=0) = ?$$

$$P(X1=1, X2=1|Y=0) = ?$$

Bayes Optimal

$$P(Y=1) = ?$$

$$P(Y=0) = ?$$

$$P(X1=1, X2=0|Y=0) = ?$$

$$P(X1=0, X2=1|Y=0) = ?$$

$$P(X1=0, X2=0|Y=0) = ?$$

$$P(X1=1, X2=1|Y=0) = ?$$

$$P(Y=1|X1=0, X2=1)$$

$$= P(X1=0, X2=1|Y=1) P(Y=1) / P(X1=0, X2=1)$$

$$= 1/2 P(X1=0, X2=1|Y=1) 4$$

$$= 2 P(X1=0, X2=1|Y=1)$$

1/2

1

XOR

X1	X2	Y
1	0	1
0	1	1
1	1	0
0	0	0

conditional independence?

Naive Bayes

$$P(Y=1) = 1/2$$

$$P(Y=0) = 1/2$$

$$P(X1=1|Y=0) = 1/2$$

$$P(X2=1|Y=0) = 1/2$$

$$P(X1=1|Y=1) = 1/2$$

$$P(X2=1|Y=1) = 1/2$$

$$P(X1=1, X2=0|Y=0) = 1/4$$

$$P(X1=0, X2=1|Y=0) = 1/4$$

$$P(X1=0, X2=0|Y=0) = 1/4$$

$$P(X1=1, X2=1|Y=0) = 1/4$$

Bayes Optimal

$$P(Y=1) = 1/2$$

$$P(Y=0) = 1/2$$

$$P(X1=1, X2=0|Y=0) = 0$$

$$P(X1=0, X2=1|Y=0) = 0$$

$$P(X1=0, X2=0|Y=0) = 1/2$$

$$P(X1=1, X2=1|Y=0) = 1/2$$

$$P(Y=1|X1=0, X2=1)$$

$$= P(X1=0, X2=1|Y=1) P(Y=1) / P(X1=0, X2=1)$$

$$= 1/2 P(X1=0, X2=1|Y=1) 4$$

$$= 2 P(X1=0, X2=1|Y=1)$$

1/2

1