

probabilistic classifies returns a probability for every lubel = posterior distribution p(Y=4/X)

"soft response"

Nocher of probabilities PC

more general than here ilassification: can recover decision function by return the most probable label

p(Y=4/X) => f(X) = arg max p(Y=4/X)

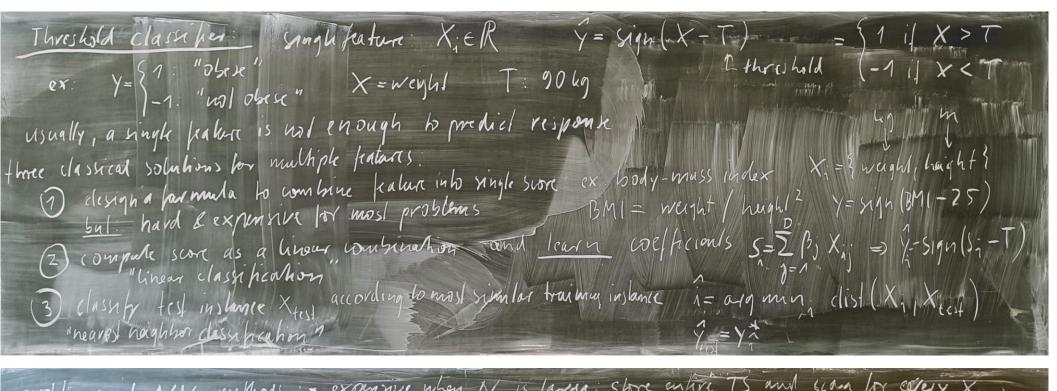
quality measured by "calibration". if p(Y=4/X)=V, then label to Should be correct V% to the laws

If achiel accuracy higher v%: "under confident" Lebech with u=Y\*

The achiel accuracy higher v%: "under confident" & dangerous, often in neural networths

lover v% "over confident" & dangerous, often in neural networths

Important: calculate confusion matrix or calibration from first Set



problems: of NN methods: - expensive when N is large store entire TS and scan for extery Xtest (fasher search exists, but scales badly with D)

- very hard to define dist (Xzed Xn) to reflect true sementic stimilarity

=> madrine learning lask "metric learning" figure out dist from training data

Bayes rule of common trained probabilities:

joint distribution of features & labels: p(X,Y) can be decomposed by chain rule of probability in two ways

find two ways

posterior

[the libroid prior p(X,Y) = p(X) p(Y|X) first measure features, then per posterior

Classification: p(Y-k|X) = \frac{p(X|Y-k)}{p(Y-k)} p(Y-k)

Bayes rule

P(X) p(X|Y) first determine response than out companies kalues

P(X) remained evidence

P(X) = \frac{p(Y-k)}{p(Y-k)} p(X-k)

Bayes Rule: p(Y=k|X) = p(X|Y=k) p(Y=k)poskerior (likelihood evidence/marginal)  $p(x) = \sum_{k=1}^{\infty} p(X|Y=k) \cdot p(Y=k)$ why is this important? ogeneral: fundamental equation for probabilite machine learning because it allows clean uncertainty handling e defines two fundamental model types
o puts model accuracy into perspective what is good or bad perf. ? turdimental madel types e cliscriminative models: learn p(Y=k|X) (LHS of Bayes)

answer question "what class" directly

relatively easy - to be direct rowle, no de bour 6 often hard to inkeptet, how model makes decisions "black box "behavior of neural networks

· generative model: learns p(Y=4) and p(X/Y=4) (RHS of Bayes) · first learn, how "the world" works. undersland mechanism, then ux this to answer question | how Observations ("phenotypes") more difficult: need powerful models

2 a bot of clasa (+) more interpretable · history: -traditional science seeks generative models: can creake synthetic data that over indistinguishable from real data - 21990: ML researchors realized that their models were too weak => field focused on discriminative models - 2012: neuval networks solved many hard discriminative tashs => field is again interested in apmenutive models (Chattaft, Midjourney)

Subfield "explainable links proclable ML" How does Bayes help to evaluate accuracy? - how good can a leasued classifier be! Del- Bayes classifier uns Bayes rule (LHS or RHS) with true prob pt Thin: No learned classifier using p coin be better than Bayes with pt

2x: p\* (Y=1/x)=0.6 then error rate less 1-60% = 40% impossible · how bad can a chisriper se? 1-p\*(Y=Y\*/X) case 1 all classes are equally probable p(Y=K)= = for all k=1,..., ( two classes p(Y=4)=0.5 worst classifier: pure questing => correct 50% of time case 2. em balanced classes, 1.9. p(Y=1)=001 p(Y=-1)=099

t mivalence does not have worst classifier always return majority lubel => 99% korrect ex: 5 reast cancer sorrening p(Y=1)=0.01 p(Y=-1)=0.99 test, eg mammo graphy p(X=test pontive | Y=1)=099 P(X=test positive/Y=-1)=0.01
"talse alaxin" Q: If a ksl is positive, should you parice p(Y=1/x=test pointie) = p(x=1/y=1)p(y=1) p(x=1|y=1)p(y=1)+p(x=1|y=-1)p(y=-1)0.19.0.01+0.01.0.99