

Resum

El rendiment de la classificació estadística està directament relacionat amb una estimació acurada de les probabilitats condicionades a la classe. Aquesta estimació es degrada críticament a mida que creix la dimensionalitat. Es poden fer diferents assumpcions per tal de simplificar l'estimació, essent la parametrització la més freqüent. Un altre apropament al problema són bàsicament tècniques de reducció de soroll. En aquest cas, qualsevol cosa que no contribueixi positivament a la classificació és considerat com a soroll. Les tècniques d'extracció i selecció de característiques s'utilitzen per a aïllar aquest soroll, preservant només les característiques rellevants. En l'espai de dimensions reduïdes és possible una estimació més robusta de les probabilitats. Les tècniques de components principals, d'anàlisi discriminant o d'escalat multidimensional fan això.

Una tercera alternativa consisteix en assumir dependències no estadístiques entre les característiques, reduint l'estimació a problemes unidimensionals. Aquest tesi aprofundeix en aquesta tercera alternativa i la connecta amb la segona des de la perspectiva següent: si la independència estadística prova una assumpció massa restrictiva, utilitza una tècnica d'extracció de característiques on aquesta hipòtesi pot ser mantinguda amb més força. Aquest enfocament també es justifica des de la perspectiva de l'error de Bayes. Des d'una perspectiva estrictament teòrica l'extracció de característiques lineal només pot incrementar aquest error, aleshores la única bona raó per aplicar una transformació lineal a les dades és la certesa que aquesta beneficiarà la precisió de l'estimació de la densitat. Els resultats en el camp de la neurociència també donen suport a aquest mètode. El cervell necessita una representació amb una redundància oculta tan petita com sigui possible. Les probabilitats d'ocurrència dels objectes i dels esdeveniments haurien de tenir una dependència estadística mínima entre ells. S'ha observat que, per a imatges naturals, la representació obtinguda mitjançant Anàlisi de Components Independents s'assembla a la representació present en el còrtex visual primari dels mamífers (V1). Aquest resultat, un cop interpretat com a una reducció de redundància natural, pot ser vista des de la nostra perspectiva com una representació que simplifica l'estimació de probabilitats.

L'Anàlisi de Components Independents és una tècnica lineal d'extracció de característiques que busca minimitzar les dependències d'alt ordre. Quan les seves assumpcions es compleixen, es poden obtenir característiques estadísticament independents a partir de les mesures originals. Adaptem l'Anàlisi de Components Independents

al problema de reconeixement estadístic de patrons de dades d'alta dimensionalitat. Això s'aconsegueix utilitzant representacions condicionals a la classe i un esquema de decisió de Bayes adaptat específicament. Degut a l'assumpció d'independència aquest esquema resulta en un modificació del classificador ingenu de Bayes. Els experiments es realitzen en un rang ampli de problemes clàssics de classificació en visió per computador, com ara el reconeixement d'objectes, aplicació de sensors remots i control de qualitat, etc.

L'inconvenient principal de les representacions condicionades a la classe és la seva incapacitat per aprendre les relacions entre les classes. Per tant també tractem el problema de la selecció de característiques dins d'aquesta representació. Un cop s'han extret les característiques s'introdueix un criteri per a seleccionar un subconjunt de característiques adequades per a la classificació. Adaptem el criteri de selecció de característiques de divergència a la nostra representació i les nostres assumpcions. L'Anàlisi de Components Independents obté components independents no Gaussians. Això fa que la divergència sigui una elecció com a mesura de la separabilitat de classes. Aquesta mesura no fa cap assumpció sobre la distribució de les dades i además es simplifica considerablement sota l'assumpció d'independència. A més a més, mostrem no es requereix una cerca exhaustiva de característiques per a obtenir el subconjunt òptim de característiques

La solució que proposem pot ser entesa des d'una perspectiva completament diferent. L'extracció de característiques independents també pot ser interpretada com a l'extracció i selecció de característiques que millor serveixen al classificador ingenu de Bayes. Extenem aquesta línia de raonament a un context no paramètric. Donat un classificador mètric com pot ser el de k veïns més propers mostrem que buscar una tècnica lineal d'extracció de característiques òptima per aquest classificador resulta en una lleugera modificació de l'anàlisi no paramètric discriminant. Aquesta connexió està justificada per resultats teòrics i pràctics. Experiments clàssics de visió per computador com ara el reconeixement de cares i de gèneres il·lustren el potencial d'aquesta tècnica.

En tots dos casos, l'enfocament estadístic paramètric i el no paramètric tractem el disseny d'un classificador de patrons com un procés integrat. Un cop les assumpcions inicials s'han complets, els diferents passos com ara l'extracció de característiques, selecció i classificació es relacionen naturalment entre ells. Tots i cadascun dels passos estan justificats teòricament i íntimament lligats amb els altres.

Abstract

Statistical classification performance is directly related with an accurate estimation of the class-conditional probabilities. For fixed sample sets, this estimate severely degrades as dimensionality increases. Different assumptions can be made such that estimation is simplified. The most frequent assumptions are those that allow the parametrization of the problem. A second approach basically consists in noise reduction techniques. Anything that does not positively contribute to classification is regarded as noise. Feature extraction and selection techniques are used to isolate this noise, preserving only relevant features. In the reduced dimensional space, a more robust parametric or nonparametric estimation of the probabilities is possible. Principal components, discriminant analysis or multidimensional scaling techniques take this approach.

A third alternative is to assume no statistical dependencies between the features, estimation being reduced to unidimensional problems. This thesis will deepen into this last approach and connect it with the second from the following perspective: if statistical independence proves a too restrictive assumption, make use of a feature extraction technique where this hypothesis can be held on stronger grounds. This approach is also justified from a Bayesian error perspective. Since, from a strictly theoretical perspective, linear feature extraction can only increase this error, the only good reason for applying a linear transform to the data is the belief that this will benefit accuracy of density estimation. Results in the field in neuroscience also support this method. The brain needs a representation with as little hidden redundancy as possible. Probabilities of occurrence of objects and events should have minimum statistical dependencies between them. Results once interpreted as natural redundancy reduction, can be seen from our perspective as a representation that simplifies probability estimation.

Independent Component Analysis (ICA) is a linear feature extraction technique that aims to minimize higher-order dependencies in the extracted features. When its assumptions are met, statistically independent features can be obtained from the original measurements. A first insight and intuitive understanding of independent component analysis is presented using this technique as a feasible model for non-rigid shape variation. We then adapt independent component analysis to the particular problem of statistical pattern recognition of high-dimensional data. This is done by means of class-conditional representations and a specifically adapted Bayesian decision

scheme. Due to the independence assumption this scheme results in a modification of the naive Bayes classifier. Experiments are made on a wide range of classical computer vision classification problems such as object recognition, remote sensing applications, quality control, etc.

The main disadvantage of class-conditional representations is that they fail to learn the relationship among classes. Therefore, we also treat the problem of feature selection within this representation. Once features are extracted a criterion is introduced for selecting a feature subset that best serves classification. We adapt the feature selection criterion of divergence to our class-conditional representation and assumptions. ICA obtains nongaussian independent components. This makes divergence an adequate choice since this class separability measure makes no assumption on the data distribution and is greatly simplified under the assumption of independence. Moreover, we show that no exhaustive search is required to obtain the optimal feature subset under our premises.

The solution we provide can be seen from a whole different perspective. The extraction of independent features can also be seen as finding a representation and then selecting those features that enhance the performance of the naive Bayes classifier. We extend this line of reasoning to a nonparametric context. Given a purely metric classifier such as the K-nearest neighbours (K-NN) we show that searching for a linear feature extraction technique optimal for this classifier results in a slight modification of nonparametric discriminant analysis. Theoretical and practical results justify this connection. Experiments in classical computer vision problems such as face and gender recognition illustrate the potential of this technique.

In both cases, the parametric statistical approach and the nonparametric metric approach, we treat the design of a pattern classifier as an integrated process. Once the initial assumptions are made the different stages such as feature extraction, selection and classification are naturally associated among each other. Each of the stages is theoretically justified and intimately linked with all the others.