

Synthetic Visual Reasoning Test Challenge

Denoncin David, Lahouel Kamel, Destagnol Kévin
Maître de stage:
Alain Trouvé

24 juin 2010

- 1 Présentation du problème
 - Exemple
- 2 Idées de bases pour l'algorithme de classification
 - Classifieur de Bayes
 - Choix des caractéristiques
 - Exemple
- 3 Problèmes techniques et pratiques, solutions
 - Problèmes d'ordres théoriques
 - Une nouvelle mesure de probabilité
 - Un nouveau critère de pertinence : l'erreur de Bayes
 - Problèmes d'ordres pratiques et techniques
- 4 Résultats
 - Test sur une base d'apprentissage de 400 images
 - Apprentissage
- 5 conclusion

- 1 Présentation du problème
 - Exemple
- 2 Idées de bases pour l'algorithme de classification
 - Classifieur de Bayes
 - Choix des caractéristiques
 - Exemple
- 3 Problèmes techniques et pratiques, solutions
 - Problèmes d'ordres théoriques
 - Une nouvelle mesure de probabilité
 - Un nouveau critère de pertinence : l'erreur de Bayes
 - Problèmes d'ordres pratiques et techniques
- 4 Résultats
 - Test sur une base d'apprentissage de 400 images
 - Apprentissage
- 5 conclusion

Exemple simple

Problem 1

Negative samples



Positive samples



Adaboost performance (1000 stumps)

Train error rate 0

Test error rate 0.467

Le Challenge

- Le Synthetic Visual Reasoning Test est un challenge concernant le machine learning et la reconnaissance de formes.
- 23 problèmes binaires où on nous présente des formes à classer selon une règle que l'on ne connaît pas a priori dont seulement 13 problèmes dévoilés à ce jour.
- En théorie, l'erreur de Bayes de chacun des problèmes est nulle, ce qui signifie qu'on peut les résoudre avec un taux d'erreur nulle, que les deux classes de formes sont disjointes.
- Les algorithmes actuels tels qu'Adaboost par exemple, bien que donnant de bons résultats en ce qui concerne certains problèmes, restent médiocres, voire mauvais pour un certain nombre d'autres.

Un deuxième exemple

Problem 21

Negative samples



Positive samples



Adaboost performance (1000 stamps)

Train error rate 0

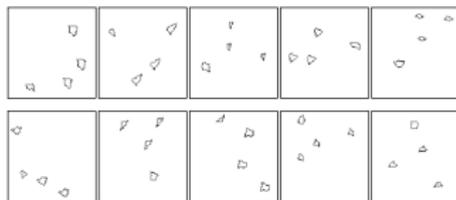
Test error rate 0.501

Figure: Un deuxième exemple

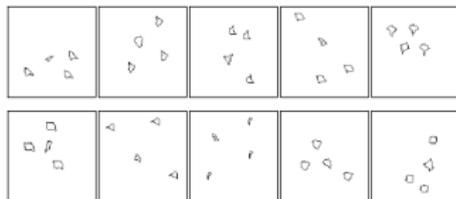
Un troisième exemple géométrique

Problem 17

Negative samples



Positive samples



Adaboost performance (1000 stamps)

Train error rate 0

Test error rate 0.388

Figure: Un troisième exemple

Approche du problème

L'idée principale est de récupérer des informations sur les tableaux pour faire un classifieur de Bayes.

- Extraire les composantes connexes pour récupérer les formes.
- Trouver des caractéristiques assez générales et prenant en compte les problèmes d'invariance

Les outils mathématiques développés et présentés lors de la soutenance de mi-stage nous ont permis de trouver des distances invariantes par similitudes, rotations et homothéties. On rappelle que nous travaillons sur la matrice de covariance (d'inertie) d'une forme et non sur la forme elle-même.

- 1 Présentation du problème
 - Exemple
- 2 Idées de bases pour l'algorithme de classification
 - Classifieur de Bayes
 - Choix des caractéristiques
 - Exemple
- 3 Problèmes techniques et pratiques, solutions
 - Problèmes d'ordres théoriques
 - Une nouvelle mesure de probabilité
 - Un nouveau critère de pertinence : l'erreur de Bayes
 - Problèmes d'ordres pratiques et techniques
- 4 Résultats
 - Test sur une base d'apprentissage de 400 images
 - Apprentissage
- 5 conclusion

Classifieur de Bayes : principe

- Un classifieur de Bayes Naïf classe un objet en faisant un test statistique qui, prend en compte un certain nombre d'informations sur l'objet pour décider à quelle classe il appartient.
- Nous choisissons ce classifieur vu qu'il est simple, robuste et souple. Nous souhaitons également, en choisissant ce classifieur, mettre l'accent sur l'importance du problème de la représentation des données dans la performance finale.

Présentation du cadre de travail

- Pour chaque problème, nous disposons d'un nombre n d'images sur lesquels on apprend. Pour chaque image, nous extrayons un nombre de caractéristiques (149).
- Nous disposons alors d'une matrice réelle où chaque ligne correspond à une image et chaque colonne à une caractéristique
- On suppose que chaque caractéristique est une variable aléatoire $X \sim \mathcal{N}(m, \sigma^2)$ dont on estime la variance et la moyenne

Rapport de vraisemblance

- $L_1(x_1, x_2, \dots, x_p) = \prod_{i=1}^p f_i^1(x_i)$
- $L_0(x_1, x_2, \dots, x_p) = \prod_{i=1}^p f_i^0(x_i)$
- Rapport de vraisemblance :

$$\frac{L_1(x_1, x_2, \dots, x_p)}{L_0(x_1, x_2, \dots, x_p)} = \frac{\prod_{i=1}^p f_i^1(x_i)}{\prod_{i=1}^p f_i^0(x_i)}$$

- La décision est prise suivant la valeur du rapport de vraisemblance

Idée de base

- On considère des relations d'équivalences sur l'ensemble des figures du tableau, et on étudie le quotient
- En pratique on fait du clustering
- L'étude du quotient se déroule en deux temps : une description intracluster et une description intercluster
- L'étude de l'image s'organise en trois niveaux de plus en plus globaux

On s'intéresse aux relations d'équivalences suivantes

- "être semblable à"
- "être une rotation de"
- "être un translaté de"
- "être homothétique à"
- "être au contact de"
- "être de même aire que"
- la relation triviale dont le seul cluster est celui comportant toutes les figures d'une image

- Niveau 0 : prise en compte d'invariances, description intracluster
- Niveau 0.5 : description intercluster
- Niveau 1 : description générale de l'image
- Niveau 2 : description "abstraite" de l'image

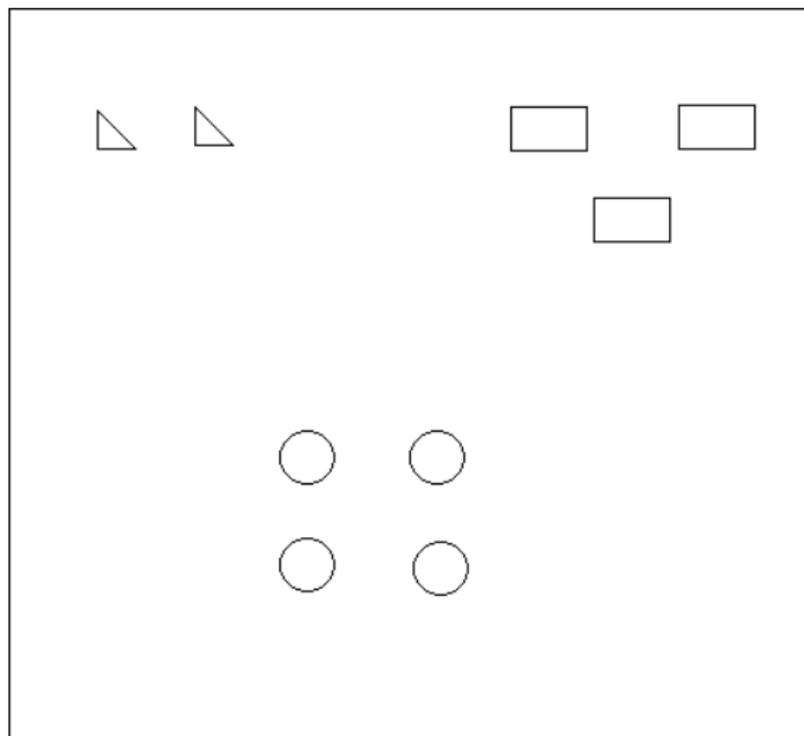


Figure: exemple de base pour le choix des caractéristiques

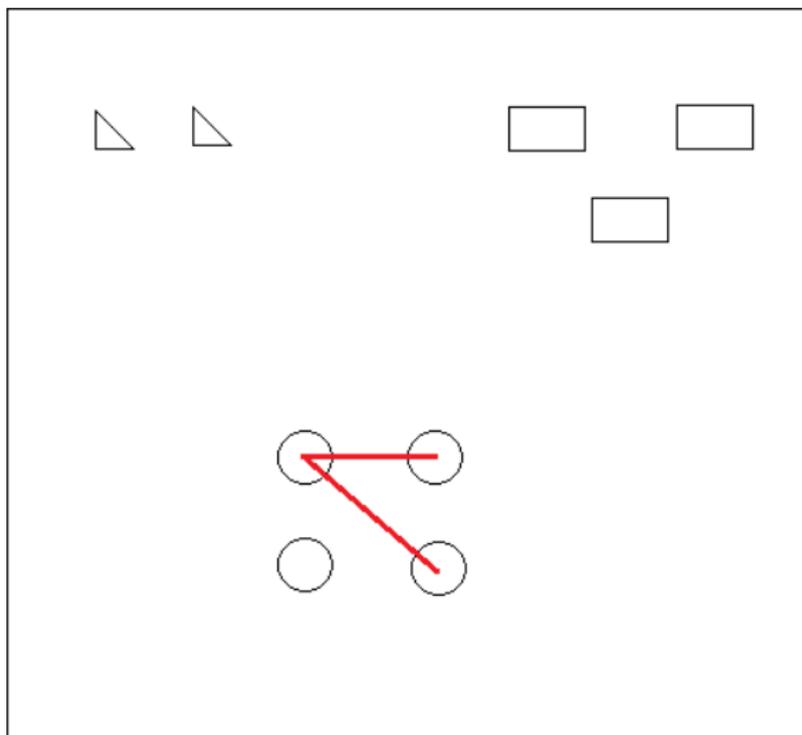


Figure: Niveau 0 : description intracluster

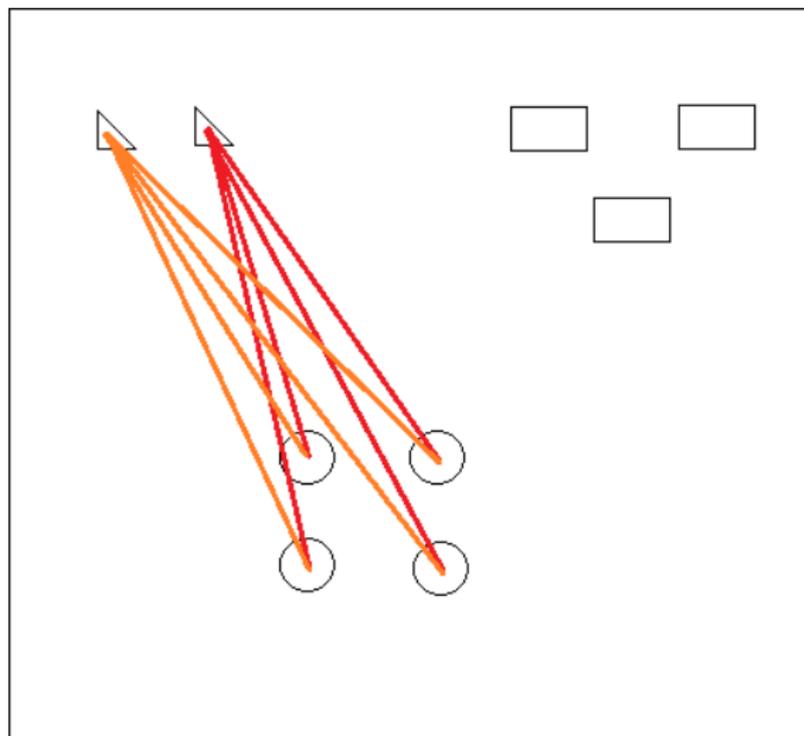


Figure: Niveau 0.5 : description intercluster

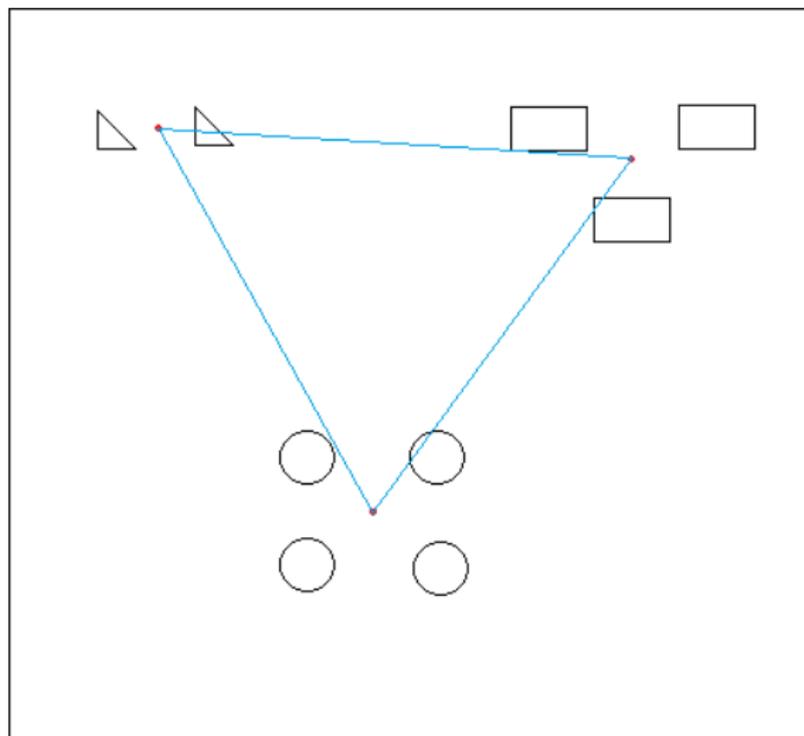


Figure: Niveau 0.5 : description intercluster

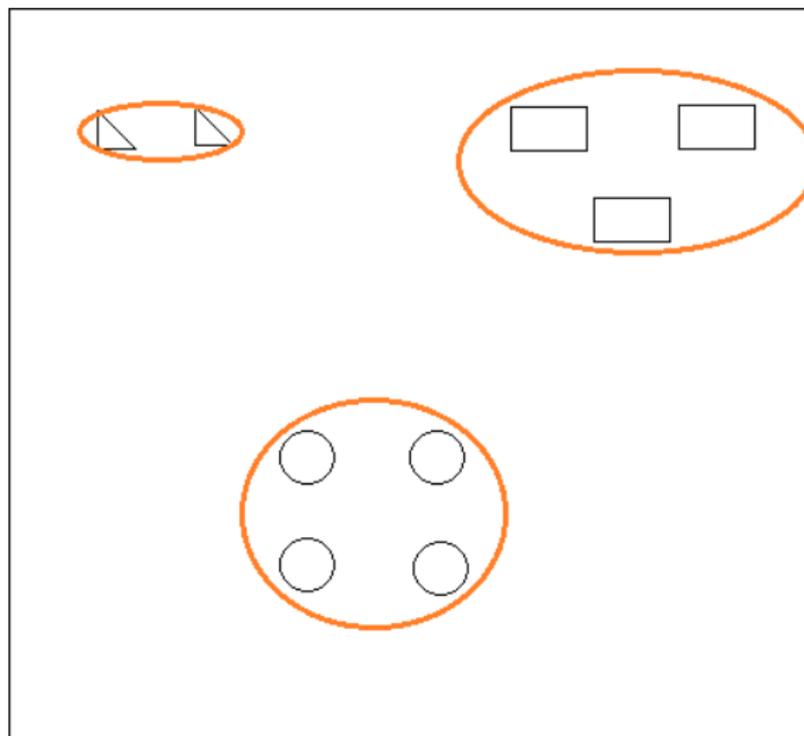


Figure: Niveau 1

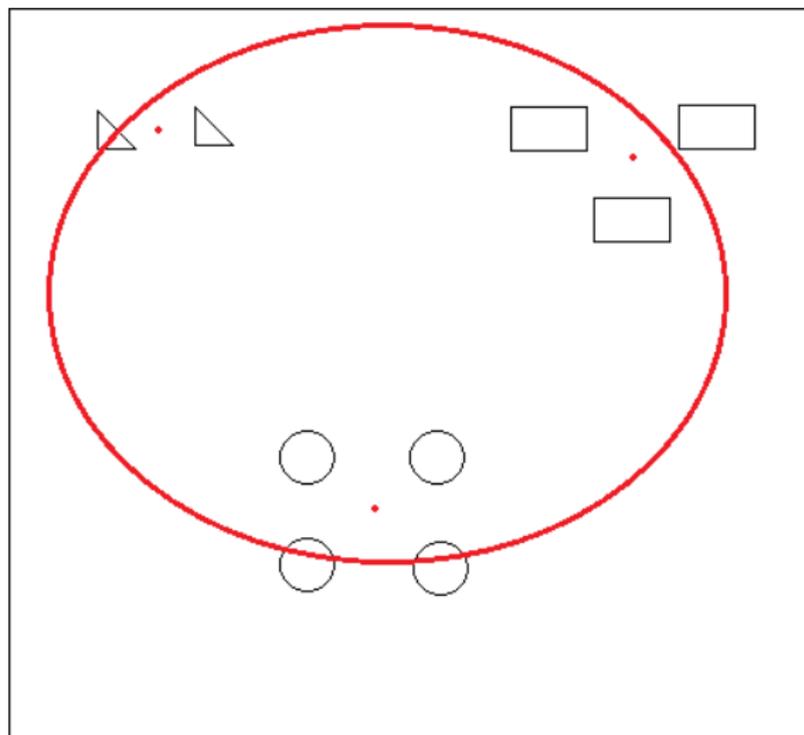


Figure: Niveau 2

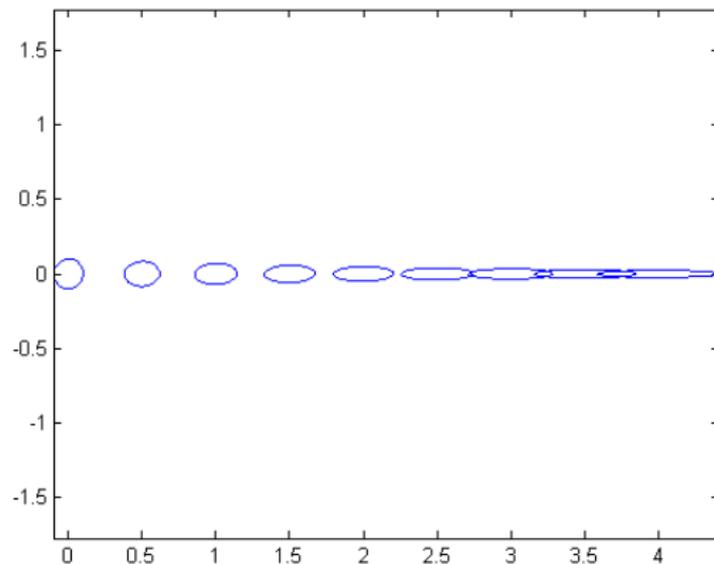


Figure: La demi-droite correspondant au quotient par les similitudes

- 1 Présentation du problème
 - Exemple
- 2 Idées de bases pour l'algorithme de classification
 - Classifieur de Bayes
 - Choix des caractéristiques
 - Exemple
- 3 **Problèmes techniques et pratiques, solutions**
 - Problèmes d'ordres théoriques
 - Une nouvelle mesure de probabilité
 - Un nouveau critère de pertinence : l'erreur de Bayes
 - Problèmes d'ordres pratiques et techniques
- 4 Résultats
 - Test sur une base d'apprentissage de 400 images
 - Apprentissage
- 5 conclusion

Inconvénients

- Nous prenons en compte toutes les caractéristiques dans le rapport de vraisemblance : certaines caractéristiques viennent "bruiter" le rapport de vraisemblance.
- Trouver un critère qui classe les caractéristiques selon leur pertinence dans le problème grâce à la phase d'apprentissage.

Un premier critère de pertinence

- L'idée est d'utiliser une sorte de distance entre 2 mesures de probabilité appelée information de Kullback.
- Pour chaque caractéristique, on regarde la distance de Kullback entre les 2 probabilités correspondant aux deux classes différentes : 0 et 1.
- Plus la distance est grande, plus les deux gaussiennes correspondant aux deux densités des deux classes sont détachées et plus la caractéristique est pertinente.

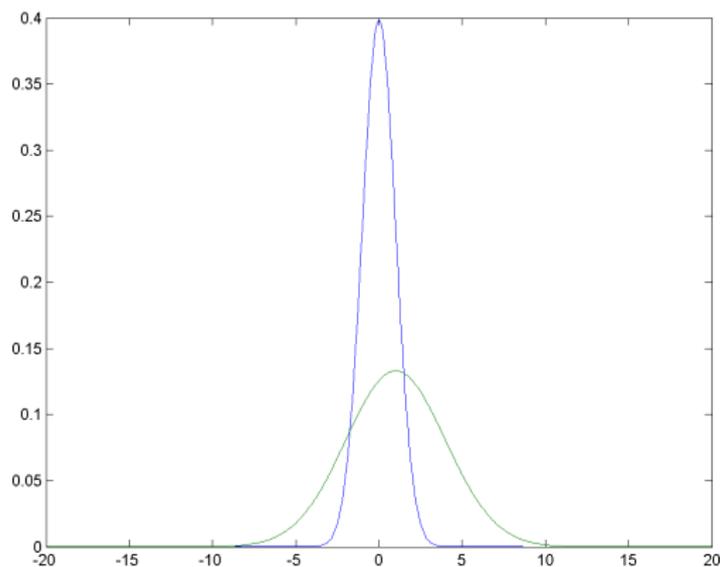


Figure: gaussiennes "détachées" une $\mathcal{N}(0, 1)$ et une $\mathcal{N}(1, 3)$ et $K(\mathcal{N}(1, 3), \mathcal{N}(0, 1)) = 5.04$

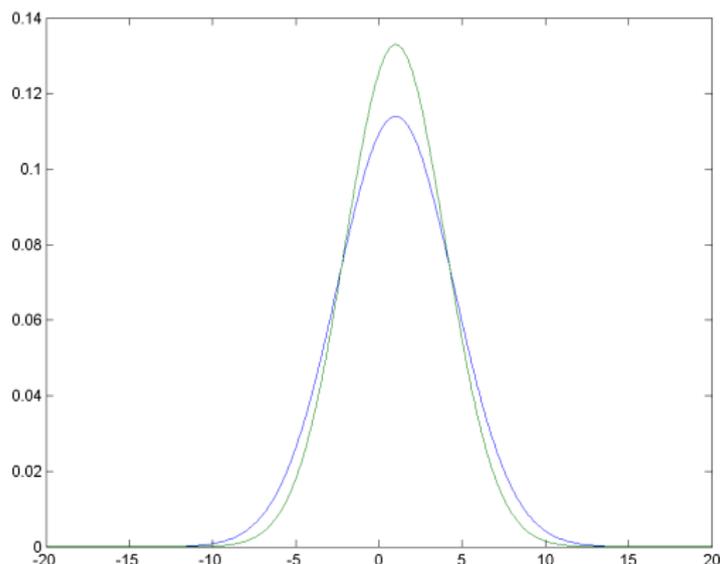


Figure: gaussiennes "moins détachées" une $\mathcal{N}(1, 3)$ et une $\mathcal{N}(1, 3.5)$ et $K(\mathcal{N}(1, 3.5), \mathcal{N}(1, 3)) = 0.2$

- La distance de KullBack se calcule facilement avec des gaussiennes
- Nous ne prenons dans le test de rapport de vraisemblance que les " n " caractéristiques qui ont les " n " distances de Kullback les plus grandes (Une bonne valeur de " n " est 10).

Autre inconvénient

- Toutes les caractéristiques ne sont pas définies pour une image donnée et nous mettons un 0 : il faut introduire un objet correspondant à "manque d'information" à la place des zéros évoqués ci-dessus.

- $\{*\}$ est attribué à la case réservée à la caractéristique si celle-ci n'est pas définie.
- Nous supposons alors que pour une caractéristique j fixée, la valeur que prend cette caractéristique si on se place dans la classe 0 suit une loi $p_j^0 \delta_* + (1 - p_j^0) \mathcal{N}(m_j^0, \sigma_j^0)$



$$m_j^0 = \frac{1}{n_{classe0 \neq *}} \sum_{i \text{ tq } x_i \neq *}^{n_{classe0}} x_i$$



$$\sigma_j^{02} = \frac{1}{n_{classe0 \neq *} - 1} \sum_{i \text{ tq } x_i \neq *}^{n_{classe0}} (x_i - m)^2$$



$$p_j^0 = \frac{1}{n_{classe0=*} + 2} \left(1 + \sum_{i \text{ tq } x_i=*}^{n_{classe0}} 1 \right)$$

- On fixe une caractéristique, on observe la "valeur" de cette caractéristique pour une image à classer.
- On fait le rapport de vraisemblance en n'utilisant que cette caractéristique.
- Si $g_1(X) \geq g_0(X)$ alors $d(X) = 1$
- Sinon $d(X) = 0$

- Notons C la variable aléatoire qui désigne la classe de l'image (0 ou 1)
- Le critère qui remplace la distance de Kullback est alors l'espérance suivante :

$$P(d(X) \neq C)$$

- En conditionnant cette espérance par rapport à X et en utilisant la formule de Bayes, on arrive à :

$$\frac{1}{2}(p_0 \wedge p_1) + \frac{1}{2} \int_{\mathbf{R}} (1 - p_0)f_0(x) \wedge (1 - p_1)f_1(x) dx$$

Estimation de l'intégrale

- $$\frac{1}{2}(\mathbf{E}(\mathbf{1}_{f_0(X) \geq f_1(X)} | C = 1) + \mathbf{E}(\mathbf{1}_{f_0(X) < f_1(X)} | C = 0))$$
- Pour estimer $\mathbf{E}(\mathbf{1}_{f_0(X) \geq f_1(X)})$ sachant que $X \sim p_1 \delta_* + (1 - p_1) \mathcal{N}(m_1, \sigma_1)$, on simule un vecteur de taille N où chacune de ses cases i est l'observation d'une variable aléatoire $Y_i = \mathbf{1}_{f_0(X_i) \geq f_1(X_i)}$ avec $X_i \sim p_1 \delta_* + (1 - p_1) \mathcal{N}(m_1, \sigma_1)$ et les X_i sont i.i.d.
- On compte alors le nombre d'apparitions de 1 dans le tableau et on divise par la taille N du tableau.

Intervalle de confiance pour notre estimateur

- Nous disposons d'un échantillon (X_1, \dots, X_N) (dans notre cas $N = 1000$) de variables aléatoires **i.i.d** suivant une **Bernoulli** de paramètre p que nous voulons estimer.
- L'inégalité de Hoeffding nous donne :

$$p \in [\overline{X}_N - \epsilon, \overline{X}_N + \epsilon]$$

avec une probabilité supérieure à $1 - 2e^{-2N\epsilon^2}$.

- Nous pouvons également construire un intervalle de confiance asymptotique avec la δ -méthode :

$$\left[\sin^2 \left(\text{Arcsin} \left(\sqrt{\overline{X}_n} - \frac{\alpha}{2\sqrt{n}} \right) \right), \sin^2 \left(\text{Arcsin} \left(\sqrt{\overline{X}_n} + \frac{\alpha}{2\sqrt{n}} \right) \right) \right]$$

Programmation des caractéristiques

- Nous avons dans une large mesure suivi les idées présentées
- Généralement le niveau 0 suffit à assurer de bons résultats en ne considérant que les translations, similitudes, contacts et aires
- Pour être performant sur les autres problèmes il faudrait programmer certains compléments

Choix des formes sur lesquelles on calcule les caractéristiques

- On ne considère que les formes qui ne sont pas trop petites par rapport à la plus grosse
- Il faut aussi un déterminant de la matrice de covariance pas trop proche de 0

Présentation du problème

Idées de bases pour l'algorithme de classification

Problèmes techniques et pratiques, solutions

Résultats

conclusion

Problèmes d'ordres théoriques

Une nouvelle mesure de probabilité

Un nouveau critère de pertinence : l'erreur de Bayes

Problèmes d'ordres pratiques et techniques



Choix de seuils

Il nous a fallu trouver des seuils pour décider si deux formes appartenait au même cluster ou non. Le plus délicat a été le seuil pour les similitudes, et nous avons utilisés les problèmes 20 et 21 pour cela.

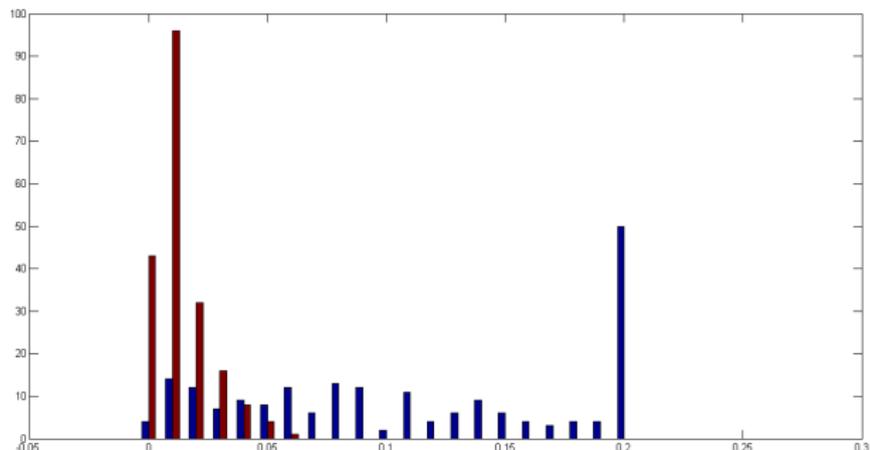


Figure: histogramme problème 20

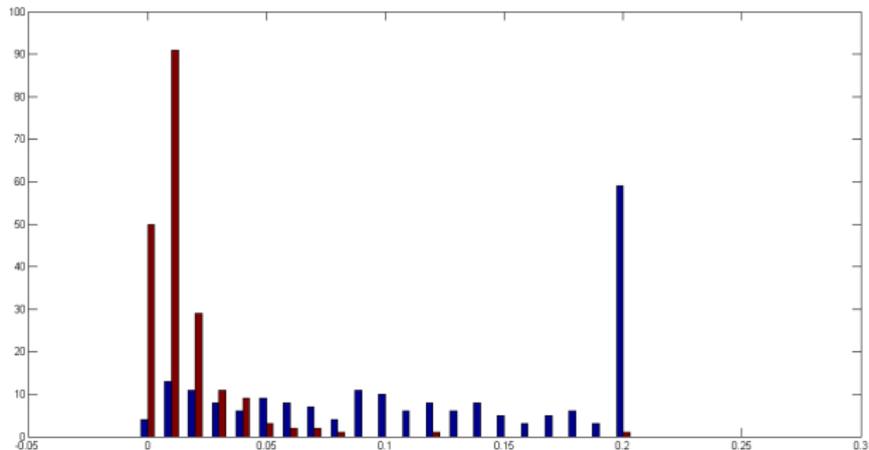


Figure: histogramme problème 21

Petits détails concernant la matrice de covariance

- Rajout de quelques points autour des barycentres lorsqu'il n'y a que deux formes dans un cluster
- On calcule la matrice de covariance de l'intérieur et il existe des images "non lues"



Figure: Image du problème 18

- 1 Présentation du problème
 - Exemple
- 2 Idées de bases pour l'algorithme de classification
 - Classifieur de Bayes
 - Choix des caractéristiques
 - Exemple
- 3 Problèmes techniques et pratiques, solutions
 - Problèmes d'ordres théoriques
 - Une nouvelle mesure de probabilité
 - Un nouveau critère de pertinence : l'erreur de Bayes
 - Problèmes d'ordres pratiques et techniques
- 4 Résultats
 - Test sur une base d'apprentissage de 400 images
 - Apprentissage
- 5 conclusion

```
Probleme 1: adaboost score 0.467 our score 0
Probleme 2: adaboost score 0.054 our score 0.0275
Probleme 3: adaboost score 0.058 our score 0.1125
Probleme 5: adaboost score 0.427 our score 0
Probleme 6: adaboost score 0.269 our score 0.0075
Probleme 8: adaboost score 0.083 our score 0.08
Probleme 11: adaboost score 0.104 our score 0.005
Probleme 12: adaboost score 0.198 our score 0
Probleme 13: adaboost score 0.399 our score 0.035
Probleme 17: adaboost score 0.388 our score 0.1
Probleme 18: adaboost score 0.067 our score 0.065
Probleme 20: adaboost score 0.482 our score 0.0825
Probleme 21: adaboost score 0.501 our score 0.1375
```

```
ans =
```

```
0.0502
```

Figure: Notre score et celui d'AdaBoost

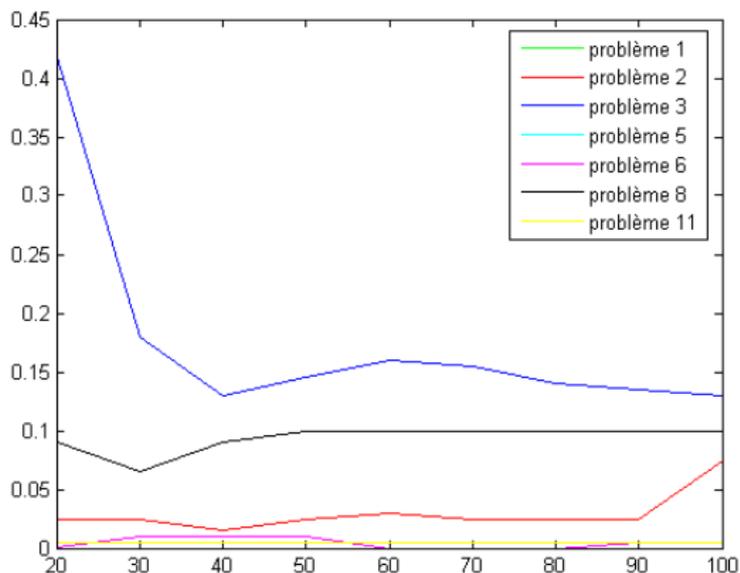


Figure: Résultats en fonction de la base d'apprentissage sur les problèmes 1,2,3,5,6,8,11

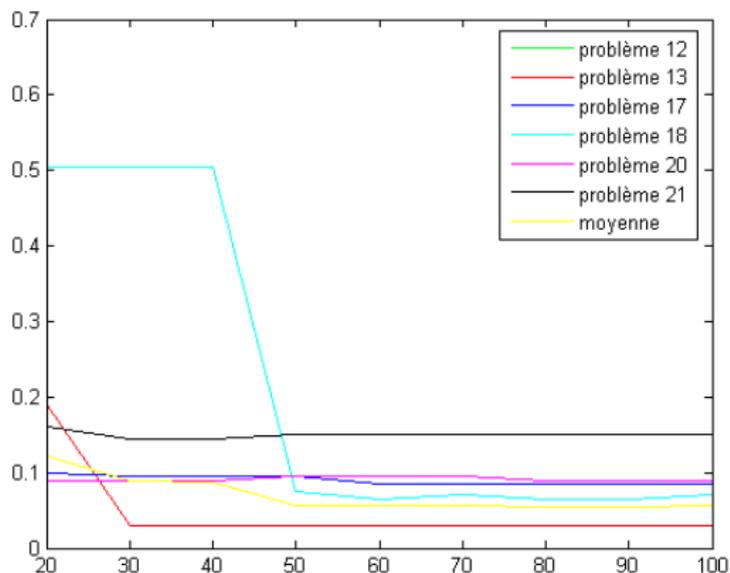


Figure: Résultats en fonction de la base d'apprentissage sur les problèmes 12,13,17,18,20,21 et l'évolution de la moyenne

- 1 Présentation du problème
 - Exemple
- 2 Idées de bases pour l'algorithme de classification
 - Classifieur de Bayes
 - Choix des caractéristiques
 - Exemple
- 3 Problèmes techniques et pratiques, solutions
 - Problèmes d'ordres théoriques
 - Une nouvelle mesure de probabilité
 - Un nouveau critère de pertinence : l'erreur de Bayes
 - Problèmes d'ordres pratiques et techniques
- 4 Résultats
 - Test sur une base d'apprentissage de 400 images
 - Apprentissage
- 5 conclusion

- Ce stage nous a initié à la recherche en groupe.
- Découvrir la géométrie différentielle et avoir une certaine avance sur le cours de M1.
- Découvrir l'univers des classifieurs.
- Appliquer quelques connaissances acquises en cours de statistiques.
- Enfin, nous sommes impatients de découvrir les performances du programme que nous avons développé sur les problèmes non dévoilés.

-  Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine.
Riemannian geometry.
Springer-Verlag, 2004.
-  Jacques Lafontaine.
Introduction aux variétés différentielles.
Presses universitaires de Grenoble, 1997.
-  Yadolah Dodge and Giuseppe Melfi.
Premiers pas en simulation.
Springer-Verlag, 2008.
-  Xavier Pennec, Pierre Fillard, and Nicholas Ayache.
A riemannian framework for tensor computing.
International Journal of Computer Vision, 2005.
-  Frédéric Helein.
Calcul différentiel et géométrie différentielle.

2002.



Synthetic visual reasoning test challenge.

www.idiap.ch/fleuret/svrt/challenge.html.



Donald Geman.

Statistical model for images.

Talk in Luminy, France, 2010.