

Comparaison de modèles et Généralisation

Soizic GESLIN
Maxime CHAMBREUIL

16 mars 2003

1 Explication

Pour ce TP, nous avons repris notre fonction discrimination, implémentée la semaine dernière, en y rajoutant quelques fonctionnalités spécifique à ce TP : Utilisation des indices d'apprentissage, de validation, calcul du meilleur coût, etc...

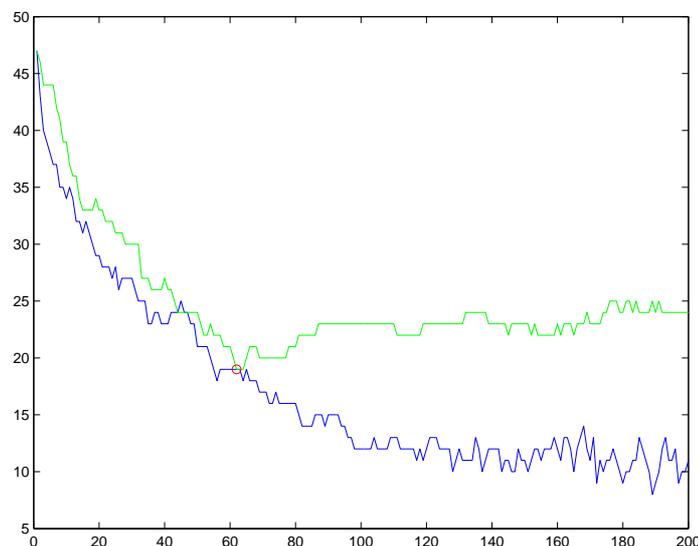
$$W, \text{Cout} = \text{discrimination}(\text{nbitemax}, \text{pas}, X, Y, \text{indicesApp}, \text{indicesVal}, \text{methode})$$

1.1 Ensemble de test

Au départ, nous avons effectué un petit réglage de nos algorithmes en sachant qu'on faisait pour tout algorithme 200 itérations. Voici nos différents résultats :

1.1.1 Perceptron

Le perceptron avec un pas de 0,00025 :

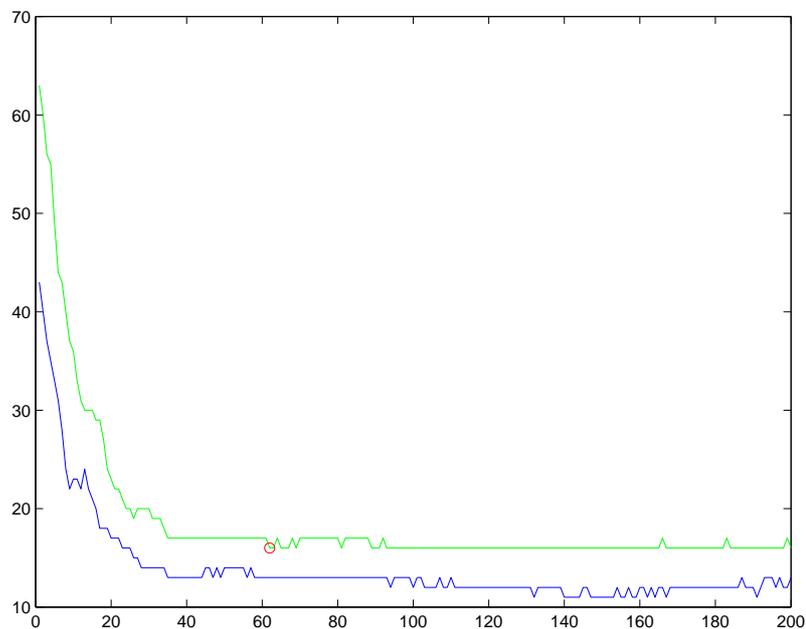


Nous obtenons 25,56 % de bonne reconnaissance et une matrice de confusion de :

		<i>Realite</i>	
		<i>M</i>	\bar{M}
<i>Prediction</i>	<i>M</i>	33	17
	\bar{M}	6	34

1.1.2 Adaline

L'adaline avec un pas de 0,001 :



Nous obtenons 18,89 % de bonne reconnaissance et une matrice de confusion de :

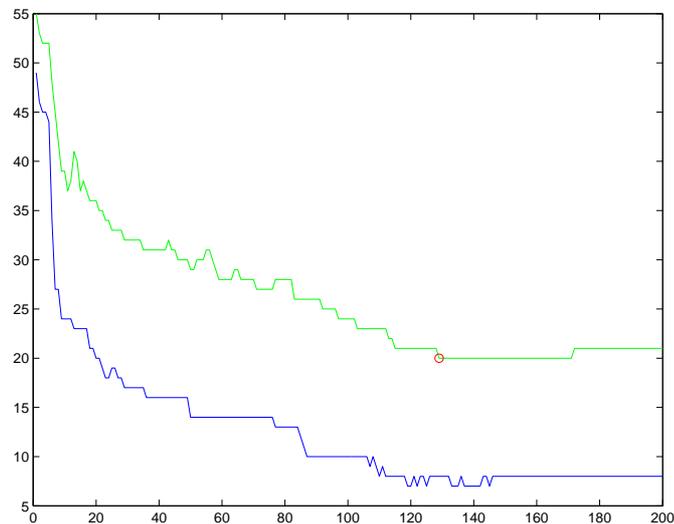
		<i>Realite</i>	
		<i>M</i>	\bar{M}
<i>Prediction</i>	<i>M</i>	37	10
	\bar{M}	7	36

1.1.3 Neurone formel

Nous obtenons 15,56 % de bonne reconnaissance et une matrice de confusion de :

		<i>Realite</i>	
		<i>M</i>	\bar{M}
<i>Prediction</i>	<i>M</i>	27	3
	\bar{M}	11	49

Le neurone formel avec un pas de 0,005 :



1.2 Validation croisée

Pour cela, nous avons juste modifié notre programme principal pour lancer des indices différents à notre fonction.

Nous avons gardé un tiers de nos données pour le test final, puis nous avons divisé le reste de notre échantillon en $L = 10$ segments, dont un pour la validation.

On obtient alors 10 modèles, que nous avons moyennés. Nous avons ensuite appliqué ce modèle sur les données de test pour calculer notre pourcentage de bonne reconnaissance et notre matrice de confusion.

1.2.1 Perceptron

Nous obtenons 17,78 % de bonne reconnaissance et une matrice de confusion de :

		<i>Realite</i>	
		M	\bar{M}
<i>Prediction</i>	M	32	7
	\bar{M}	9	42

1.2.2 Adaline

Nous obtenons 12,22 % de bonne reconnaissance et une matrice de confusion de :

		<i>Realite</i>	
		<i>M</i>	\bar{M}
<i>Prediction</i>	<i>M</i>	32	4
	\bar{M}	7	47

1.2.3 Neurone formel

Nous obtenons 14,44 % de bonne reconnaissance et une matrice de confusion de :

		<i>Realite</i>	
		<i>M</i>	\bar{M}
<i>Prediction</i>	<i>M</i>	39	5
	\bar{M}	8	38

1.3 Régularisation

J'ai choisi de considérer l'algorithme du neurone formel, parce que c'est l'algorithme qui me semble le plus robuste et constant dans ses différents résultats.

Régulariser a consisté pour nous à rajouter un terme au calcul du cout de notre fonction :

$$cout = \frac{\sum abs(sign(X_{indicesApp} * W_{App}) - Y(indicesApp))}{2} + \lambda * (norm(W_{App}))^2$$

Voici les différentes valeurs de notre modèle et le pourcentage pour différents lambda :

λ	0	0.2	0.5	0.7	1	2	5	10
<i>W</i>	-0,33	-0,04	0,18	0,18	-0,27	-0,27	-0,27	-0,04
	-0,53	-0,52	-0,46	-0,58	-0,62	-0,58	-0,59	-0,42
	-1,02	-0,73	-0,56	-0,68	-1,11	-0,95	-0,95	-1,18
	0,05	0,13	-0,75	0,40	0,16	-0,27	-0,27	-0,23
	-0,43	-0,41	0,09	0,21	-0,15	0,00	0,00	-0,18
	-0,14	0,33	0,20	0,41	-0,18	0,13	0,13	0,00
	-0,46	-0,50	-0,35	-0,10	-0,04	-0,04	-0,04	-0,34
	0,18	-,016	0,32	-0,22	0,02	0,08	0,08	-0,35
	-0,26	-0,72	-0,54	-0,41	-0,53	-0,45	-0,45	-0,46
	-0,94	-0,62	-0,24	-0,20	-0,48	-0,45	-0,45	-0,93
	-0,14	-0,64	-0,75	-0,86	-0,25	-0,48	-0,48	-0,50
	-1,07	-1,01	-1,17	-1,24	-1,47	-0,79	-0,79	-0,95
	-0,64	-0,46	-0,62	-0,38	-0,92	-0,77	-0,77	-0,52
	0,13	0,68	0,15	0,59	0,16	0,19	0,18	0,01
Mal classe :	22,22%	26,67%	20%	14,44%	20%	20%	20%	22,22%

2 Interprétation

Il est assez difficile d'interpréter les résultats en raison de la grande dimension des données : Le pas et le nombre d'itération pour chaque algorithme sont très instable l'un par rapport à l'autre. C'est pour cette raison que nous décidés de choisir un nombre d'itération fixe pour chaque et ainsi les comparer beaucoup plus facilement.

2.1 Notion de sur-apprentissage

On peut remarquer sur les premières courbes, que la courbe verte (cout en validation) remonte après le point rouge (cout minimum). On illustre de cette façon la notion de sur-apprentissage : On en sait déjà assez sur le modèle, en apprendre plus ne sert à rien. C'est comme les mathématiques : une fois qu'on connaît les chiffres, on connaît tous les nombres. Il n'y a pas besoin de tous les apprendre par coeur...

2.2 Généralisation

En généralisant, on va pouvoir comparer les différentes méthodes. Mais il existe plusieurs méthodes de généralisation. Au vue des résultats, on peut dire que la validation croisée généralise mieux que l'ensemble de test. C'est logique, étant donné qu'on combine plusieurs fois la même méthode sur les mêmes données : un même individu a servi 10 fois lors de la validation, alors qu'il a servi une seule fois dans l'ensemble de test.

Durant nos tests sur la validation croisée, nous avons eu comme première intention de ne pas prendre le cout moyenné mais de prendre le cout pondéré : nous voulions donner plus de poids à un modèle qui donnait de meilleur résultat, en multipliant chaque modèle par son pourcentage de bien classé par exemple. Mais il fallait prendre un ensemble de test pour lui appliquer le modèle moyenné et ne pas prendre un ensemble de test pour chaque segment.

2.3 Régularisation

La régularisation doit permettre de simplifier les composantes du modèle. C'est comme de tirer sur les rennes d'un cheval au galop : on retient le modèle pour éviter qu'il se ne complique. La subtilité réside dans le fait de maintenir le pourcentage de bonne reconnaissance (maintenir le cheval à la même vitesse).

Nous avons considéré qu'une valeur de composante était plus simple lorsque ses chiffres significatifs diminuaient. En considérant aussi le pourcentage, nous en avons déduit que 2 était notre meilleure valeur de λ . En-dessous, les composantes sont compliquées. Au delà, le pourcentage augmente.

3 Conclusion

Ce TP nous a permis de nous rendre compte à quel point les réglages d'un algorithme sont importants. Après pas mal de temps, nous avons pu déterminer leurs paramètres de prédilection. Après cette recherche sur les différents pas des boucles, nous espérons aller un petit plus vite à l'avenir dans une recherche similaire.

Ce TP était intéressant de part les notions qu'il abordait : sur-apprentissage, généralisation, etc... A chaque fois, nous avons pu faire le parallèle avec notre propre méthode de travail : apprendre par coeur du vocabulaire en langue, juste connaître les grandes lignes d'un cours, retenir les formules ou savoir les redémontrer, etc... Nous en concluons qu'un apprentissage régulier et jamais dans le même ordre ne nous garantira pas la validation, mais nous en assurera les meilleurs chances.