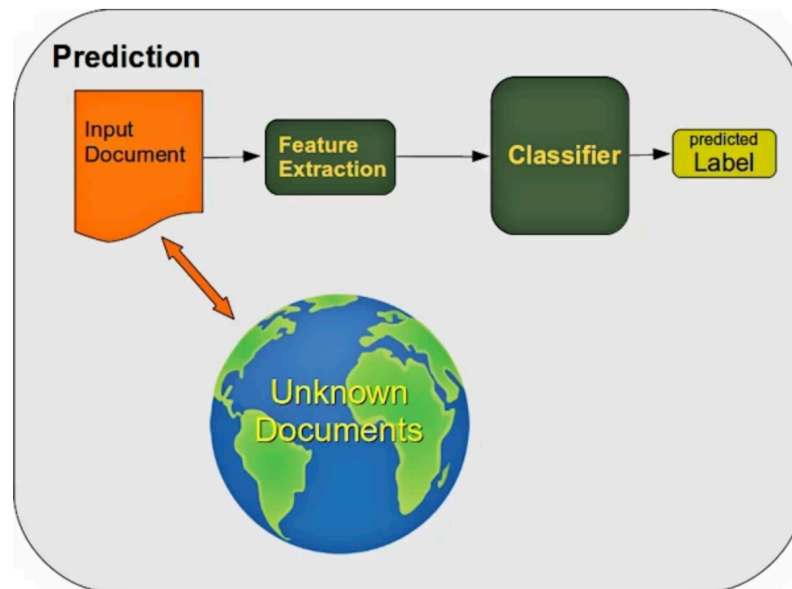


Terminologie de l'apprentissage automatique

Classifieur

Un programme ou une fonction qui fait correspondre des instances non étiquetées à des classes est appelé un classificateur.



Une matrice de confusion, également appelée table de contingence ou matrice d'erreur, est utilisée pour visualiser les performances d'un classificateur.

Les colonnes de la matrice représentent les instances des classes prédites et les lignes représentent les instances de la classe réelle. (Remarque : cela peut aussi être l'inverse).

Dans le cas d'une classification binaire, le tableau comporte 2 lignes et 2 colonnes.

Exemple :

Confusion Matrix		Predicted classes	
		male	female
Actual classes	male	42	8
	female	18	32

Cela signifie que le classificateur a correctement prédit une personne masculine dans 42 cas et qu'il a prédit à tort que 8 instances masculines étaient des femmes. Il a correctement prédit une personne féminine dans 32 cas. 18 cas ont été prédits à tort comme étant des hommes au lieu de femmes.

Précision (accuracy)

La précision est une mesure statistique qui est définie comme le quotient des prédictions correctes faites par un classificateur divisé par la somme des prédictions faites par le classificateur.

Dans notre exemple précédent, le classificateur a correctement prédit 42 instances masculines et 32 instances féminines.

Par conséquent, la précision peut être calculée par :

$$acc = \frac{42+32}{42+8+18+32} = 0,72$$

Supposons que nous ayons un classifieur qui prédit toujours **femme**. Nous avons une précision de 50 % dans ce cas.

		Predicted classes	
		male	female
Actual classes	male	0	50
	female	0	50

Nous allons démontrer le paradoxe de la précision.

Un classificateur de reconnaissance de spam est décrit par la matrice de confusion suivante :

		Predicted classes	
		spam	ham
Actual classes	spam	4	1
	ham	4	91

La précision de ce classificateur est de $\frac{4+91}{100} = 95\%$.

Le classificateur suivant prédit uniquement "jambon" et a la même précision.

		Predicted classes	
		spam	ham
Actual classes	spam	0	5
	ham	0	95

La précision de ce classificateur est de 95%, même s'il n'est pas capable de reconnaître le moindre spam.

Précision et rappel

		Predicted classes	
		negative	positive
Actual classes	negative	TN	FP
	positive	FN	TP

$$accuracy = \frac{T_P + T_N}{T_P + F_P + T_N + F_N}$$

$$precision = \frac{T_P}{T_P + F_P}$$

$$recall = \frac{T_P}{T_P + F_N}$$

Apprentissage supervisé

Le programme d'apprentissage automatique reçoit à la fois les données d'entrée et l'étiquetage correspondant. Cela signifie que les données d'apprentissage doivent être étiquetées au préalable par un être humain.

Apprentissage non supervisé

Aucune étiquette n'est fournie à l'algorithme d'apprentissage. L'algorithme doit déterminer le regroupement des données d'entrée.

Apprentissage par renforcement

Un programme informatique interagit dynamiquement avec son environnement. Cela signifie que le programme reçoit un retour positif et/ou négatif pour améliorer ses performances.

Entrée []: