

Fiche d'auto-évaluation 12

May 27, 2020

1. Quelle est la différence entre des technique de sélection de caractéristiques (feature extraction) et l'analyse en composantes principales (Principal Component Analysis)
2. Comment peut-on utiliser l'analyse en composantes principales (Principal Component Analysis) pour la détection des anomalies?
3. Quand on applique l'analyse en composantes principales, comment choisit-on le nombre de composantes à retenir?
4. Écrivez la formule de décomposition en valeurs singuliers (Singular Value Decomposition) d'une matrice de donnée \mathbf{A} . Écrivez les propriétés de chaque terme de la formule.
5. Si on fait une décomposition en valeurs singuliers (Singular Value Decomposition) sur une matrice de données et on met en graphique le nombre de composantes vs. la variance capturée par ces composantes, quel est la forme du graphe que vous obtenez? Dessinez-le, approximativement.
6. Prenez une matrice de données \mathbf{A} et appliquez une décomposition en valeurs singuliers (Singular Value Decomposition). Supposons que $\mathbf{A}^{(k)}$ est la matrice réduite, obtenue en retenant seulement k composantes. A quoi est égale la différence entre \mathbf{A} et $\mathbf{A}^{(k)}$, mesurée par la norme de Froebenius, c'est à dire $\|\mathbf{A} - \mathbf{A}^{(k)}\|_F$?
7. Qu'est-ce que Elastic Search? Un modèle d'apprentissage automatique? Un algorithme? Quelque chose d'autre?
8. Dans la suite Elastic on retrouve Logstash, Elastic Search et Kibana. Quel est le rôle de chacun?
9. Qu'est l'indexation qui Elastic Search implémente? À quoi sert elle?
10. Quel composant lit les données brutes et les transforme en format json, lisible par Elastic Search? Kibana ou Logstash?
11. Qu'est-ce que Netflow? Donnez un exemple du type d'informations qu'on peut obtenir à travers Netflow.

12. Supposons qu'à travers l'interface Kibana on voit dans notre base de donnée une étiquette "France" comme pays d'origine d'un paquet. Parmi les composantes de la suite Elastic, qui a créé cette information textuelle?
13. What is Permutation Importance and how is it computed?
14. Que prédit le Kaplan-Meier estimator? Est-il utile pour classifier une type d'anomalie? Ou pour avoir une idée de la probabilité de faillite d'une machine?
15. Qu'est-ce que la fonction de survie (survival function) $S(t) = \Pr(T > t)$ signifie?
16. La fonction de survie $S(t) = \Pr(T > t)$ est elle monotone? Si oui, décroissante ou croissante? Quelle est sa valeur pour $t = 0$? Et pour $t \rightarrow +\infty$?
17. Quels sont les différents modèles que nous avons vu pour estimer la probabilité de survie? Sont ils des modèles supervisés ou non-supervisés?