

ДПО МЛ 5 - снижение размерности

***Обязательный вопрос**

1. Электронная почта *

2. Ваше имя и фамилия *

3. Имеется корпус из трехсот текстов, размер словаря которого составляет 6000 слов. К матрице векторов документов текста применили снижение размерности путем сингулярного разложения с количеством компонент, равным 3. Какова будет размерность сжатой матрицы? * 1 балл

Ответ должен состоять из двух чисел, разделенных запятой.

4. Выберите все **методы снижения размерности**, при помощи которых можно извлечь значимые кластеры из данных. * 1 балл

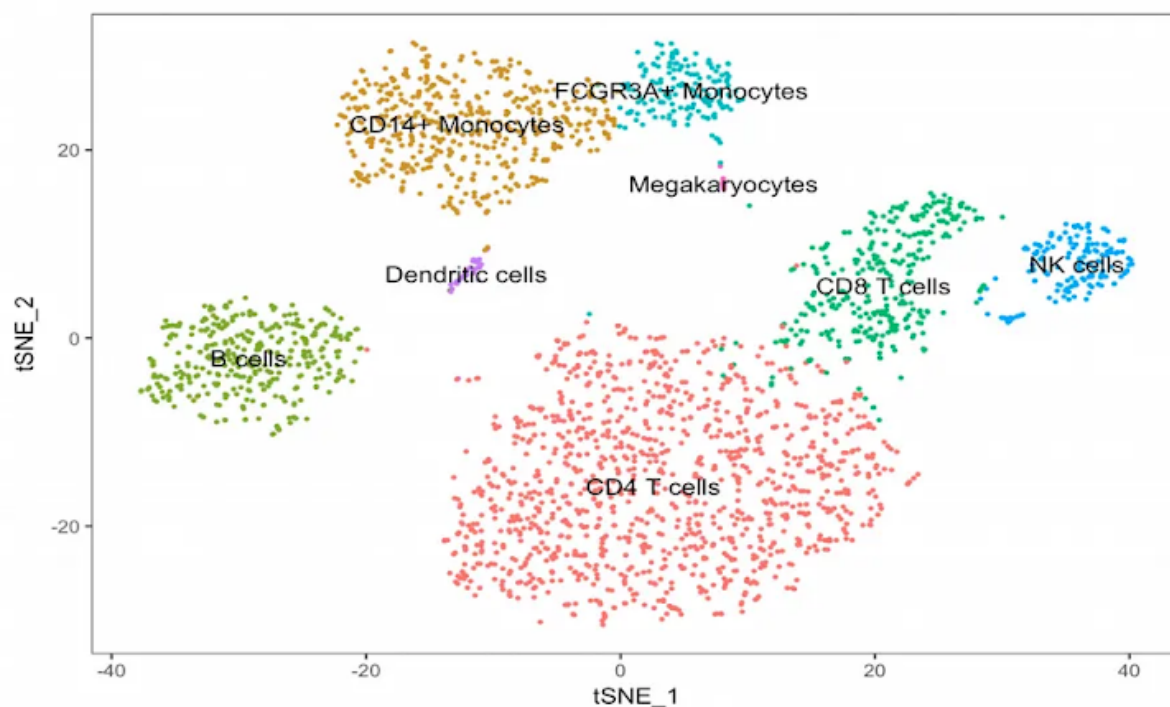
Отметьте все подходящие варианты.

- ☐ Latent Dirichlet Allocation
☐ K Means
☐ t-SNE

5. Данная визуализация была создана при помощи tSNE (автор - Николай Осколков: <https://towardsdatascience.com/how-to-tune-hyperparameters-of-tsne-7c0596a18868>).

* 1 балл

Можем ли мы с уверенностью утверждать, что кластеры NK cells и B cells находятся далеко друг от друга в реальном (не сжатом) пространстве признаков?



Отметьте только один овал.

☐ Да

☐ Нет

6. Выберите все верные утверждения о перплексии. *

1 балл

Отметьте все подходящие варианты.

- ☐ Большие значения перплексии на тестовых данных означают лучшее качество модели
- ☐ Перплексия - важный гиперпараметр для tSNE
- ☐ При помощи перплексии можно оценить качество латентного размещения Дирихле
- ☐ При помощи перплексии можно оценить качество любого метода снижения размерности

7. Имеется корпус из 5 текстов, размер словаря которого - 100 слов. * 1 балл
Сколько дисперсии объяснят **все** компоненты сингулярного разложения **совокупно**, если n_components равно размеру словаря?

Отметьте только один овал.

- ☐ 100%
- ☐ 36%
- ☐ Не хватает данных для ответа

8. Данный код должен решать задачу уменьшения размерности корпуса * 1 балл
текстов X. Считайте, что импорты и разделение на тренировочную и тестовую выборки прошли в предыдущих клетках. В какой из строк вы заметили ошибку?

```
1 X_train = [...]  
2 y_train = [1, 2, 3...]  
3  
4 tfidf = TfidfVectorizer()  
5 X_train_transformed = tfidf.fit_transform(X_train)  
6  
7 lda = LinearDiscriminantAnalysis(n_components=10)  
8 lda.fit(X_train_transformed)
```

Отметьте только один овал.

- ☐ 5
- ☐ 7
- ☐ 8
- ☐ Ошибка в нехватке кода, а не в конкретной строке

Компания Google не имеет никакого отношения к этому контенту.

Google Формы

