1. В чем суть линейных классификаторов? Какие методы относятся к

линейным классификаторам?

Линейные классификаторы - это алгоритмы машинного обучения, которые строят разделяющую гиперплоскость в пространстве признаков для разделения объектов разных классов. Они основаны на предположении о линейной разделимости данных. Некоторые из методов линейных классификаторов включают логистическую регрессию, метод опорных векторов (SVM) с линейным ядром и линейные дискриминантные анализы.

2. Что такое регуляризация?

Регуляризация - это метод контроля за сложностью модели путем добавления штрафа за большие веса или параметры модели. Это позволяет предотвратить переобучение и улучшить обобщающую способность модели.

3. Объясните назначение L1 и L2-регуляризации?

L1 и L2-регуляризация представляют собой два подхода к регуляризации. L1-регуляризация (Lasso) добавляет штраф к функции потерь, равный сумме абсолютных значений весов модели. L2-регуляризация (Ridge) добавляет квадратичный штраф к функции потерь, равный сумме квадратов весов модели.

4. Объясните влияние параметров С и гамма на регуляризацию модели?

Параметры C и гамма влияют на регуляризацию модели в методе опорных векторов (SVM). Параметр C контролирует штраф за неправильную классификацию объектов. Большее значение C приводит к меньшему штрафу, что может привести к переобучению. Гамма определяет влияние одного обучающего примера и варьирует ширину ядра SVM: маленькое значение гаммы означает широкий радиус ядра, а большое значение - узкий радиус.

5. Какой метод помогает подобрать лучшую комбинацию параметров?

Метод перекрестной проверки (кросс-валидации) помогает подобрать лучшую комбинацию параметров модели путем оценки ее производительности на различных подмножествах данных. Например, GridSearchCV в библиотеке scikit-learn может использоваться для систематического поиска лучших параметров модели путем проверки всех комбинаций значений параметров.

6. Что такое специфичность и чувствительность модели?

Специфичность (specificity) модели - это доля истинно отрицательных случаев, которые были правильно классифицированы. Чувствительность (sensitivity) модели - это доля истинно положительных случаев, которые были правильно классифицированы.

7. Как по ROC-кривой можно сравнить модели. Что означает AUC в ROC-

кривой?

При сравнении моделей по ROC-кривой, оценивается их способность различать между классами. ROC-кривая показывает отношение между чувствительностью и специфичностью модели при различных порогах классификации. AUC (Area Under Curve) в ROC-кривой представляет собой площадь под кривой ROC и представляет собой общую производительность модели: чем выше AUC, тем лучше модель способна различать между классами.