**1. \*\*Суть линейных классификаторов\*\*:**

Линейные классификаторы относятся к алгоритмам машинного обучения, которые строят разделяющую гиперплоскость в пространстве признаков для разделения объектов на различные классы. Основная идея заключается в том, чтобы найти оптимальную линейную функцию (гиперплоскость), которая максимально хорошо разделяет объекты разных классов.

Примеры линейных классификаторов:

- Логистическая регрессия

- Метод опорных векторов (SVM)

- Линейный дискриминантный анализ (LDA)

- Персептрон

**2. \*\*Регуляризация\*\*:**

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

**3. \*\*Назначение L1 и L2-регуляризации\*\*:**

- \*\*L1-регуляризация (Lasso)\*\*: Добавляет штраф к функции потерь, пропорциональный сумме абсолютных значений весов модели. Это приводит к разреженности весов, что означает, что некоторые признаки могут быть полностью исключены из модели.

- \*\*L2-регуляризация (Ridge)\*\*: Добавляет штраф к функции потерь, пропорциональный сумме квадратов весов модели. Это способствует уменьшению весов, но они редко становятся нулевыми, поэтому все признаки остаются в модели.

**4. \*\*Влияние параметров C и гамма на регуляризацию модели\*\*:**

- \*\*Параметр C влияет на силу регуляризации\*\*: Большие значения C уменьшают силу регуляризации, что позволяет модели настраиваться на обучающие данные более точно, но может привести к переобучению. Маленькие значения C увеличивают силу регуляризации, что делает модель более устойчивой к переобучению.

- \*\*Параметр гамма (только для SVM)\*\*: Он влияет на радиус ядра SVM. Большие значения гамма приводят к более сложным моделям, что может привести к переобучению, особенно при недостаточном количестве обучающих данных. Маленькие значения гамма приводят к более простым моделям, но могут не хватать для точного разделения классов в случае сложных данных.

**5. \*\*Метод, который помогает подобрать лучшую комбинацию параметров\*\*:**

Для подбора лучших параметров модели часто используют кросс-валидацию с поиском по сетке (Grid Search) или случайный поиск параметров (Random Search).

**6. \*\*Специфичность и чувствительность модели\*\*:**

- \*\*Специфичность (Specificity)\*\*: Показатель способности модели правильно классифицировать отрицательные случаи (истинно отрицательные) из всех отрицательных случаев. Формально это отношение true negative (TN) к сумме true negative (TN) и false positive (FP).

- \*\*Чувствительность (Sensitivity)\*\*: Показатель способности модели правильно классифицировать положительные случаи (истинно положительные) из всех положительных случаев. Формально это отношение true positive (TP) к сумме true positive (TP) и false negative (FN).

**7. \*\*Сравнение моделей по ROC-кривой. AUC в ROC-кривой\*\*:**

- ROC-кривая (Receiver Operating Characteristic) позволяет сравнивать качество моделей на основе их способности разделять классы путем изменения порога классификации. Она представляет собой график зависимости чувствительности от специфичности при различных значениях порога.

- AUC (Area Under Curve) - это площадь под ROC-кривой. Чем больше значение AUC, тем лучше модель. AUC равен вероятности того, что случайно выбранный положительный экземпляр будет классифицирован выше, чем случайно выбранный отрицательный экземпляр. Если AUC равен 1, модель идеальна; если AUC равен 0,5, это означает случайное угадывание.