

Вопрос 1:

Объясните, почему токенизация является первым и критически важным шагом в любой NLP системе. Приведите пример, когда неправильная токенизация может привести к неправильным результатам.

Ключевые концепции: токенизация, морфемы, регулярные выражения

Вопрос 2:

Когда вы будете использовать простую токенизацию по пробелам, а когда BPE? Объясните преимущества и недостатки каждого подхода на примере технического текста с множеством неизвестных слов.

Ключевые концепции: BPE, подсловная токенизация, словарь, размер словаря

Вопрос 3:

Напишите регулярное выражение для извлечения всех email адресов из текста. Объясните каждую часть вашего regex и почему вы выбрали именно такой шаблон.

Ключевые концепции: регулярные выражения, звезда Клини

Вопрос 4:

Почему Unicode и UTF-8 критически важны для обработки многоязычных текстов? Может ли система, построенная только для английского языка (используя ASCII), работать для русского или китайского текста?

Ключевые концепции: Unicode, UTF-8, кодирование символов, многоязычность

Вопрос 5:

Почему информация о том, как были собраны данные, как они были аннотированы и кто их аннотировал, важна для оценки качества NLP системы? Приведите пример, где эта информация повлияла бы на мою оценку модели.

Ключевые концепции: метаданные, демография, процесс сбора данных, аннотирование, bias

Вопрос 6:

Объясните цепное правило вероятности для последовательности из трех слов. Как марковское предположение упрощает вычисления? Приведите пример с реальными вероятностями.

Ключевые концепции: цепное правило, совместная вероятность, условная вероятность, марковское предположение

Вопрос 7:

Почему биграммная модель ($N=2$) часто лучше чем триграммная модель ($N=3$), несмотря на то что триграммы имеют больше контекста? Какой компромисс здесь существует?

Ключевые концепции: N-граммы, биграммы, триграммы, данные тренировки, разреженность данных

Вопрос 8:

Что такое перплексия и почему это хорошая метрика для языковых моделей?

Интерпретируйте следующие результаты: модель А имеет перплексию 50, модель В имеет перплексию 200. Какая лучше и почему?

Ключевые концепции: перплексия, оценка качества, языковая модель, метрики

Вопрос 9:

Объясните, почему нужны отдельные наборы для обучения, разработки и тестирования. Что произойдет, если я буду тестировать модель на данных, которые были в обучающем наборе? Как это называется?

Ключевые концепции: обучающий набор, тестовый набор, набор разработки, загрязнение данных, переобучение

Вопрос 10:

Как логистическая регрессия решает задачу бинарной классификации? Почему сигмоид преобразует выход в вероятность?

Ключевые концепции: логистическая регрессия, сигмоид, вероятность, линейная граница, веса

Вопрос 11:

Объясните, почему кросс-энтропийная потеря подходит для классификации. Как она штрафует неправильные предсказания? Сравните ее с квадратичной потерей (MSE).

Ключевые концепции: кросс-энтропийная потеря, отрицательная логарифмическая вероятность, функция потерь

Вопрос 12:

Объясните алгоритм градиентного спуска пошагово. Что происходит с весами на каждой итерации? Почему выбор скорости обучения критичен?

Ключевые концепции: градиентный спуск, градиент, скорость обучения, итерация, сходимость

Вопрос 13:

Вы построили классификатор спама с точностью (ассурасу) 95%. Но когда вы смотрите на матрицу ошибок, вы видите, что он предсказывает "не-спам" для 95% всех примеров. Что произошло? Какие метрики вы должны использовать вместо ассурасы?

Ключевые концепции: матрица ошибок, точность, полнота, F1-мера, дисбаланс классов

Вопрос 14:

Почему модель может переобучаться даже на правильно разделенных данных? Как L1 и L2 регуляризация предотвращают переобучение? Когда вы используете одну вместо другой?

Ключевые концепции: регуляризация, L1, L2, переобучение, штраф, веса

Вопрос 15:

Объясните дистрибутивную гипотезу. Почему слова с похожими контекстами имеют похожие значения? Приведите примеры (например, рецепты, спортивные новости).

Ключевые концепции: дистрибутивная гипотеза, контекст, семантика, сходство слов, контекстное окно

Вопрос 16:

Объясните алгоритм Skip-gram в Word2Vec. Почему вместо обучения на полном наборе признаков мы используем контекстное окно? Как это помогает моделям масштабироваться?

Ключевые концепции: Word2Vec, Skip-gram, CBOW, контекстное окно, отрицательная выборка

Вопрос 17:

Объясните косинусное сходство и евклидово расстояние. Почему косинусное сходство часто предпочтительнее для эмбедингов слов? Приведите пример с реальными числами.

Ключевые концепции: косинусное сходство, евклидово расстояние, метрики, нормализация

Вопрос 18:

Как можно использовать эмбединги для обнаружения синонимов, антонимов и отношений гиперонимия/гипонимия? Приведите пример арифметики встраиваний: king - man + woman \approx ?

Ключевые концепции: синонимы, антонимы, гиперонимия, гипонимия, аналогии, векторная арифметика

Вопрос 19:

Как эмбединги слов используются в качестве входа для нейронных сетей при классификации текста? Как они отличаются от ручного извлечения признаков?

Ключевые концепции: встраивания как признаки, обучение представлений, ручные признаки, нейронные сети

Вопрос 20:

Объясните архитектуру простой трехслойной нейросети для классификации текста: входной слой → скрытый слой → выходной слой. Как посчитать количество параметров сети?

Ключевые концепции: слой, входной слой, скрытый слой, выходной слой, размер сети

Вопрос 21:

Сравните ReLU, tanh и сигмоид как функции активации. Какие проблемы решает каждая? Почему ReLU стал стандартом в современных сетях?

Ключевые концепции: функция активации, ReLU, tanh, сигмоид, исчезающие градиенты

Вопрос 22:

Покажите пошагово, как узел нейросети вычисляет свой выход. Объясните роль весов, смещения и функции активации.

Ключевые концепции: узел, вес, смещение, взвешенная сумма, функция активации

Вопрос 23:

Объясните алгоритм обратного распространения. Как градиент вычисляется от выходного слоя к входному? Почему важна дифференцируемость функций активации?

Ключевые концепции: обратное распространение, градиент, цепное правило, дифференцируемость

Вопрос 24:

Как эмбединги слов используются в качестве входа для нейросети? Объясните, почему умножение на one-hot вектор эквивалентно индексированию в embedding matrix.

Ключевые концепции: embedding matrix, one-hot вектор, индексирование, эмбединги

Вопрос 25:

Объясните методы mean-pooling и max-pooling. Когда используют каждый? Как они преобразуют последовательность встраиваний в один вектор?

Ключевые концепции: pooling, mean-pooling, max-pooling, агрегирование, порядок слов

Вопрос 26:

Объясните три проблемы, которые могут возникнуть при обучении нейросетей: переобучение, недообучение и исчезающие градиенты. Как решать каждую?

Ключевые концепции: переобучение, недообучение, исчезающие градиенты, регуляризация, ранняя остановка, функции активации

Вопрос 27:

Объясните, как предварительное обучение больших языковых моделей работает. Почему обучение на миллиардах параметров и петабайтах текста дает лучшие результаты, чем меньшие модели?

Ключевые концепции: предварительное обучение, масштаб, большие языковые модели, обучение без разметки

Вопрос 28:

Объясните разницу между архитектурами декодер, энкодер и энкодер-декодер. Приведите примеры моделей для каждой архитектуры и задач, которые они решают.

Ключевые концепции: декодер, энкодер, энкодер-декодер, двунаправленность

Вопрос 29:

Объясните разницу между жадным декодированием и выборкой при генерации текста. Какие преимущества и недостатки у каждого подхода?

Ключевые концепции: жадное декодирование, выборка, температура, детерминизм, разнообразие

Ключевые концепции: top-k выборка, nucleus выборка, вероятностная масса, усечение

Вопрос 30:

Объясните общую архитектуру трансформера. Какие компоненты используются? Как они работают вместе для обработки текста?

Ключевые концепции: трансформер, входной слой, трансформер блок, выходной слой, стек слоев

Вопрос 31:

Что такое самовнимание и как оно работает? Объясните роль Query, Key и Value. Почему это позволяет модели фокусироваться на релевантных частях текста?

Ключевые концепции: самовнимание, Query, Key, Value, матрица внимания, релевантность

Вопрос 32:

Объясните многоголовое внимание. Почему использование нескольких голов лучше чем один большой head? Что изучает каждый head?

Ключевые концепции: многоголовое внимание, головы внимания, параллельные вычисления, разные типы внимания

Вопрос 33:

Почему трансформер нужно информировать о позиции каждого токена? Как это делается? Почему нельзя просто прибавить номер позиции?

Ключевые концепции: позиционное кодирование, позиция, относительное расстояние, sin/cos функции

Вопрос 34:

Объясните концепцию "scaling laws" для больших языковых моделей. Как производительность зависит от размера модели, размера данных и объема вычислений?

Ключевые концепции: scaling laws, размер модели, размер данных, вычисления, параметры