

# Программа экзамена по машинному обучению

ml-mipt basic, весна 2020

1. Постановка задач машинного обучения, их классификация, примеры.
2. Метрики качества классификации: accuracy, balanced accuracy, precision, recall, f1-score, ROC-AUC, multiclass extensions.
3. Метрики качества регрессии: MSE, MAE, R2, другие варианты.
4. Метод максимального правдоподобия (MLE).
5. Наивный байесовский классификатор.
6. Метод ближайшего соседа (kNN).
7. Линейная регрессия. Постановка задачи линейной регрессии. Аналитическое решение МНК, оптимальность оценки, теорема Гаусса-Маркова (формулировка).  
Градиентное решение задачи линейной регрессии.
8. Регуляризация: L1 и L2. Свойства, вероятностная интерпретация.
9. Логистическая регрессия.  
Эквивалентность решений полученных методом максимального правдоподобия и минимизации логистической функции потерь.
10. Стратегии мультиклассовой классификации, их свойства.
11. Метод опорных векторов (SVM), ядра (kernel trick).
12. Алгоритм PCA. Связь с SVD, теорема Эккарта-Янга (формулировка), связь с дисперсией.  
Практическое применение.
13. Этапы построения модели: тренировка, валидация, тестирование. Роль каждого этапа.  
Переобучение.
14. Стратегии валидации. Утечки (leaks).
15. Bias-variance tradeoff.
16. Понятие информации, информационной энтропии.  
Критерии информативности: энтропийный, Джини.
17. Жадный алгоритм построения дерева. Стандарты для построения деревьев.
18. Ансамблирование. Процедура bootstrap. Бэггинг. Метод случайных подпространств.
19. Random Forest, другие леса.
20. Бустинг. Градиентный бустинг.
21. Матричные вычисления. Матричное дифференцирование.  
Производные основных функций:  $a^T x$ ,  $Ax$ .
22. Backpropagation, chain rule.
23. Neural network concept. Fully-Connected layer (FC).  
Logistic regression as simple NN.  
XOR problem.
24. Losses for NNs: logistic loss, cross-entropy.

25. Activation functions, their impact on the network, computational complexity. Softmax and LogSoftmax activations, numerical stability.
26. Методы оптимизации для обучения нейронных сетей.  
Градиентный спуск (GD). Стохастический градиентный спуск (SGD).  
Adaptive gradient methods: adagrad, adamax, adadelta. RMSprop, Adam.
27. Регуляризация нейронных сетей: Dropout, Batch Normalization. Разница в поведении при обучении и предсказании.
28. Vanilla Recursive NN cell. Backpropagation through RNN.  
Vanishing gradient problem. Potential solutions.
29. LSTM/GRU, memory concept, gates ideas.
30. Matrix convolution. Convolutional layer, backpropagation through it. Hyperparameters of Convs.  
1x1 convolutions, comparison to FC layers. Max/Average Pooling.
31. Архитектуры сетей для классификации изображений. Основные достижения каждой.
32. Работа с категориальными признаками и пропущенными значениями.  
Mean and label encoding.

*Данные вопросы не попали в основную лекционную программу не будут оцениваться на экзамене, но на них их полезно уметь отвечать.*

- *Text mining: Bag of Words, TF-IDF.*
- *Word2vec. Skip-gram, negative sampling, treating idioms as "words".  
Word2vec as matrix factorization (optional).*
- *Геометрические методы машинного обучения и кластеризация. IsoMap, LLE, DBSCAN, k-means, t-SNE*

## Теоретический минимум

1. *Постановка задач обучения с учителем (supervised learning).*
2. *Задачи обучения без учителя. Назвать хотя бы две.*
3. *Что означает свойство i.i.d.?*
4. *Основная идея наивного Байесовского классификатора. В чём его наивность?*
5. *Запишите формулы для модели линейной регрессии и для среднеквадратичной ошибки.*
6. *Запишите формулу для одного шага градиентного спуска. Как модифицировать градиентный спуск для очень большой выборки?*
7. *Что такое правдоподобие, метод максимального правдоподобия? Является ли правдоподобие вероятностью?*
8. *Что такое кросс-валидация? На что влияет количество блоков в кросс-валидации?*
9. *Что такое переобучение и недообучение? Как их можно детектировать?*
10. *Чем гиперпараметры отличаются от параметров? Что является параметрами и гиперпараметрами в линейных моделях и в решающих деревьях?*

11. Что такое регуляризация? Чем на практике отличается L1-регуляризация от L2?
12. Учитывается ли коэффициент сдвига  $w_0$  в регуляризаторе? Почему?
13. Почему линейные модели рекомендуется применять к выборке с нормированными значениями признаков?
14. Запишите формулу для линейной модели классификации. Что такое отступ?
15. Что такое точность и полнота? Почему нужно учитывать их вместе?
16. В задаче бинарной классификации доля одного класса составляют 95% выборки. Какие метрики разумно использовать для оценки работы модели? почему?
17. Что такое ROC-AUC? Как построить ROC-кривую?
18. Запишите функционал логистической регрессии. Как он связан с методом максимума правдоподобия?
19. Идея метода опорных векторов (в случае разделимой выборки).
20. Опишите жадный алгоритм обучения решающего дерева.
21. Почему с помощью решающего дерева можно достичь нулевой ошибки на обучающей выборке без повторяющихся объектов?
22. Если в лист дерева попали объекты разных классов, то какие предсказания нужно выдавать в этом листе? Почему?
23. Какое предсказание нужно выдавать в листе дерева в задаче регрессии если мы минимизируем MSE? а в случае MAE?
24. Что такое bagging?
25. Что такое случайный лес? Чем он отличается от бэггинга над решающими деревьями?
26. Как в градиентном бустинге обучаются базовые алгоритмы?
27. Зачем нужен backprop, что такое производная вектора по вектору?
28. Опишите принцип работы свёрточного слоя (CNN).
29. В чем недостатки полносвязных нейронных сетей какая мотивация к использованию свёрточных?
30. Опишите принцип работы базового рекуррентного слоя (RNN).
31. Что такое dropout?
32. Как dropout и batch normalization меняют свое поведение при эксплуатации модели (в режиме inference)?
33. Запишите постановку задачи в методе главных компонент.
34. Как работает метод k-means?