

## Определения и формулировки

1. Выпуклое множество.
2. Выпуклая функция.
3.  $\mu$ -сильно выпуклая функция.
4. Необходимые и достаточные условия локального экстремума.
5. Принцип Ферма.
6. Метод градиентного спуска.
7. Наискорейший спуск.
8.  $L$ -гладкость.
9. Липшицева парабола для  $L$ -гладкой функции.
10. Условие Поляка-Лоясевича.
11. Характер сходимости градиентного спуска для гладких невыпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
12. Характер сходимости градиентного спуска для гладких выпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
13. Характер сходимости градиентного спуска для гладких сильно выпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
14. Сходимость градиентного спуска для сильно выпуклых квадратичных функций. Оптимальные гиперпараметры.
15. Нижние оценки для минимизации гладких выпуклых функций с помощью методов первого порядка в терминах  $\Omega$  от числа итераций.
16. Нижние оценки для минимизации гладких сильно выпуклых (PL) функций с помощью методов первого порядка в терминах  $\Omega$  от числа итераций.
17. Отличие ускоренной линейной сходимости от неускоренной.
18. Связь методов первого порядка и многочленов при минимизации квадратичной функции.
19. Ключевое (минимаксное) свойство полиномов Чебышёва.
20. Метод тяжёлого шарика (Поляка) для квадратичной функции. Характер сходимости. Оптимальные гиперпараметры.
21. Формулировка ускоренного градиентного метода Нестерова для выпуклых гладких функций.
22. Характер сходимости ускоренного градиентного метода Нестерова для выпуклых гладких функций в терминах  $\mathcal{O}$  от числа итераций.
23. Формулировка ускоренного градиентного метода Нестерова для сильно выпуклых гладких функций.
24. Характер сходимости ускоренного градиентного метода Нестерова для сильно выпуклых гладких функций в терминах  $\mathcal{O}$  от числа итераций.
25. Метод Ньютона.
26. Сходимость метода Ньютона для квадратичной функции.
27. Характер сходимости метода Ньютона для сильно выпуклых гладких функций — куда и как сходится.
28. Демпфированный метод Ньютона.
29. Метод BFGS.
30. Основная идея L-BFGS.
31. Проекция.
32. Достаточное условие существования проекции точки на множество.
33. Достаточное условие единственности проекции точки на множество.
34. Метод проекции градиента.
35. Критерий проекции точки на выпуклое множество (неравенство Бурбаки-Чейни-Гольдштейна).
36. Проекция как нестягивающий оператор.
37. Характер сходимости метода проекции градиента для гладких выпуклых функций в терминах  $\mathcal{O}$

- от числа итераций.
38. Характер сходимости метода проекции градиента для гладких сильно выпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
  39. Метод Франк-Вульфа.
  40. Характер сходимости метода Франк-Вульфа для гладких выпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
  41. Характер сходимости метода Франк-Вульфа для гладких сильно выпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
  42. Субградиент. Субдифференциал.
  43. Субградиентный метод.
  44. Характер сходимости субградиентного метода для негладких выпуклых Липшицевых функций в терминах  $\mathcal{O}$  от числа итераций.
  45. Характер сходимости субградиентного метода для негладких сильно выпуклых Липшицевых функций в терминах  $\mathcal{O}$  от числа итераций.
  46. Стратегия выбора шага для субградиентного метода в случае выпуклой Липшицевой функции.
  47. Нижние оценки для негладкой выпуклой оптимизации с помощью методов первого порядка в терминах  $\Omega$  от числа итераций.
  48. Нижние оценки для негладкой сильно выпуклой оптимизации с помощью методов первого порядка в терминах  $\Omega$  от числа итераций.
  49. Проксимальный оператор.
  50. Оператор проекции как частный случай проксимального оператора.
  51. Характер сходимости проксимального градиентного метода для гладких выпуклых функций  $f$  в терминах  $\mathcal{O}$  от числа итераций.
  52. Характер сходимости проксимального градиентного метода для гладких сильно выпуклых функций  $f$  в терминах  $\mathcal{O}$  от числа итераций.
  53. Аналитическое выражение для  $\text{prox}_{\lambda\|x\|_1}$ .
  54. Аналитическое выражение для  $\text{prox}_{(\mu/2)\|x\|_2^2}$ .
  55. Проксимальный оператор как нерастягивающий оператор.
  56. Характер сходимости ускоренного проксимального градиентного метода для гладких выпуклых функций  $f$  в терминах  $\mathcal{O}$  от числа итераций.
  57. Метод стохастического градиентного спуска.
  58. Идея мини-батча. Эпоха.
  59. Характер сходимости стохастического градиентного спуска для гладких выпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
  60. Характер сходимости стохастического градиентного спуска для гладких PL-функций в терминах  $\mathcal{O}$  от числа итераций.
  61. Характер работы стохастического градиентного спуска с постоянным шагом для гладких PL-функций.
  62. Метод SVRG.
  63. Метод SAG.
  64. Метод Adam.
  65. Метод AdamW.
  66. Как сравниваются методы в AlgoPerf Benchmark.
  67. Сопряжённая (двойственная) норма.
  68. Linear Minimization Oracle (LMO).
  69. Связь LMO с сопряжённой нормой.
  70. Неевклидов градиентный спуск.
  71. Неевклидов нормированный градиентный спуск.
  72. Неевклидов градиентный спуск в случае, когда аргумент — матрица.
  73. Неевклидов нормированный спуск в случае, когда аргумент — матрица.

74. Полярный фактор матрицы.
75. LMO для спектральной нормы.
76. Формулировка метода нормированного спуска в норме Фробениуса.
77. Формулировка метода нормированного спуска в поэлементной  $\ell_\infty$ -норме.
78. Формулировка метода нормированного спуска в спектральной норме.
79. Метод Muon.
80. Итерации Ньютона-Шульца для вычисления полярного фактора.
81. Идея проекции функции потерь нейронной сети на прямую, плоскость.
82. Grokking.
83. Double Descent.
84. Идея gradient checkpointing.
85. Идея аккумуляции градиентов.
86. Зачем увеличивать батч при обучении больших моделей.
87. Warmup.
88. Правила масштабирования learning rate для больших батчей (линейное и корневое).
89. Сколько памяти занимают веса крупнейшей открытой модели (DeepSeek V4-Pro,  $\approx 1,6$  трлн параметров) в форматах FP32, FP16, BF16.
90. Mixed Precision Training.
91. LoRA.
92. Преимущество cooldown-фазы по сравнению с cosine decay.
93. Метод двойственного градиентного подъёма.
94. Связь константы сильной выпуклости  $f$  и гладкости  $f^*$ .
95. Сопряжённая функция.
96. Связь субградиентов функции  $f$  и сопряжённой  $f^*$ .
97. Идея dual decomposition.
98. Метод двойственной декомпозиции для линейных ограничений-неравенств.
99. Метод модифицированной функции Лагранжа.
100. Метод ADMM.
101. Формулировка задачи линейных наименьших квадратов с  $\ell_1$ -регуляризацией в форме ADMM.
102. Формулировка задачи поиска точки на пересечении двух выпуклых множеств в форме ADMM.
103. Дифференциальное уравнение градиентного потока.
104. Характер сходимости траектории градиентного потока для выпуклых функций в терминах  $\mathcal{O}$  от времени  $t$ .
105. Характер сходимости траектории градиентного потока для PL-функций в терминах  $\mathcal{O}$  от времени  $t$ .
106. Дифференциальное уравнение Нестеровского ускоренного градиентного потока.
107. Характер сходимости траектории Нестеровского ускоренного градиентного потока для выпуклых функций в терминах  $\mathcal{O}$  от времени  $t$ .
108. Стохастический градиентный поток.
109. Уравнение Фоккера-Планка.
110. Проклятие размерности методов нулевого порядка.
111. Эволюционные стратегии.
112. Поиск по сетке и случайный поиск.
113. Метод имитации отжига.
114. Метод роя частиц (PSO).
115. Муравьиный алгоритм (ACO).
116. SMA-ES.
117. Идея AlphaEvolve.

## Теоремы с доказательствами

1. Теорема сходимости градиентного спуска для гладких выпуклых функций.
2. Теорема сходимости градиентного спуска для гладких PL-функций.
3. Теорема сходимости градиентного спуска для сильно выпуклых квадратичных функций. Оптимальные гиперпараметры.
4. Теорема о сходимости метода Ньютона для сильно выпуклых функций с липшицевым гессианом.
5. Вывод формул обновления (прямого и обратного) гессиана для SR-1 и DFP.
6. Вывод формул обновления (прямого и обратного) гессиана для BFGS.
7. Вывод и пояснение формулы L-BFGS.
8. Теорема о сходимости метода проекции градиента для выпуклой гладкой функции.
9. Теорема о сходимости метода проекции градиента для сильно выпуклой гладкой функции.
10. Теорема о сходимости метода Франк-Вульфа для выпуклой гладкой функции.
11. Теорема о нижней оценке для методов, обращающихся к допустимому множеству только через оракул линейной минимизации (LMO).
12. Теорема о сходимости субградиентного метода для выпуклых Липшицевых функций. Стратегии выбора шага.
13. Теорема о сходимости субградиентного метода для сильно выпуклых Липшицевых функций.
14. Теорема о нижней оценке для негладкой выпуклой оптимизации методами первого порядка.
15. Теорема о сходимости проксимального градиентного метода для выпуклой гладкой функции.
16. Теорема о сходимости проксимального градиентного метода для сильно выпуклой гладкой функции.
17. Теорема о сходимости SGD в гладком PL-случае.
18. Теорема о связи сильной выпуклости функции  $f$  и липшицевости градиента сопряжённой  $f^*$ .
19. Теорема сходимости траектории градиентного потока для выпуклых и PL-функций.
20. Теорема о сходимости траектории Нестеревского ускоренного градиентного потока для выпуклых функций.