Определения и формулировки

1. Положительно определённая матрица.
2. Евклидова норма вектора.
3. Неравенство треугольника для нормы.
4. p-норма вектора.
5. Как выглядит единичный шар в p - норме на плоскости для $p = 1, 2, \infty$?
6. Норма Фробениуса для матрицы.
7. Спектральная норма матрицы.
8. Скалярное произведение двух векторов.
9. Скалярное произведение двух матриц, согласованное с нормой Фробениуса.
10. Собственные значения матрицы. Спектр матрицы.
11. Связь спектра матрицы и её определенности.
12. Спектральное разложение матрицы.
13. Сингулярное разложение матрицы.
14. Связь определителя и собственных чисел для квадратной матрицы.
15. Связь следа и собственных чисел для квадратной матрицы.
16. Линейная сходимость последовательности.
17. Сублинейная сходимость последовательности.
18. Сверхлинейная сходимость последовательности.
19. Квадратичная сходимость последовательности.
20. Тест корней для определения скорости сходимости последовательности.
21. Тест отношений для определения скорости сходимости последовательности.
22. Унимодальная функция.
23. Метод дихотомии.
24. Метод золотого сечения.
25. Метод параболической интерполяции.
26. Условие достаточного убывания для неточного линейного поиска.
27. Условия Гольдштейна для неточного линейного поиска.
28. Условие ограничения на кривизну для неточного линейного поиска.
29. Градиент функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.
30. Гессиан функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.
31. Якобиан функции $f(x) : \mathbb{R}^n \to \mathbb{R}^m$.
32. Формула для аппроксимации Тейлора первого порядка $f^I_{x_0}(x)$ функции $f(x) : \mathbb{R}^n \to \mathbb{R}$ в точке x_0.
33. Формула для аппроксимации Тейлора второго порядка $f^{II}_{x_0}(x)$ функции $f(x) : \mathbb{R}^n \to \mathbb{R}$ в точке x_0.
34. Связь дифференциала функции df и градиента ∇f для функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.
35. Связь второго дифференциала функции $d^2 f$ и гессиана $\nabla^2 f$ для функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.
36. Формула для приближенного вычисления производной функции $f(x) : \mathbb{R}^n \to \mathbb{R}$ по k-ой координате с помощью метода конечных разностей.
37. Пусть $f = f(x_1(t), \ldots, x_n(t))$. Формула для вычисления $\frac{\partial f}{\partial t}$ через $\frac{\partial x_i}{\partial t}$ (Forward chain rule).
38. Пусть L - функция, возвращающая скаляр, а v_k - функция, возвращающая вектор $x \in \mathbb{R}^t$. Формула для вычисления $\frac{\partial L}{\partial v_k}$ через $\frac{\partial L}{\partial x_i}$ (Backward chain rule).
39. Идея Хатчинсона для оценки следа матрицы с помощью matvec операций.
40. Афинное множество. Афинная комбинация. Афинная оболочка.
41. Выпуклое множество. Выпуклая комбинация. Выпуклая оболочка.
42. Конус. Выпуклый конус. Коническая комбинация. Коническая оболочка.
43. Внутренность множества.
44. Относительная внутренность множества.

45. Сумма Минковского.
46. Любые 2 операции с множествами, сохраняющие выпуклость.
47. Выпуклая функция.
48. Строго выпуклая функция.
49. Надграфик функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.
50. Множество подуровней функции $f(x) : \mathbb{R}^n \to \mathbb{R}$.
51. Дифференциальный критерий выпуклости первого порядка.
52. Дифференциальный критерий выпуклости второго порядка.
53. Связь выпуклости функции и её надграфика.
54. μ-сильно выпуклая функция.
55. Дифференциальный критерий сильной выпуклости первого порядка.
56. Дифференциальный критерий сильной выпуклости второго порядка.
57. Любые 2 операции с функциями, сохраняющие выпуклость.
58. Теорема Тейлора.
59. Необходимые условия локального экстремума.
60. Достаточные условия локального экстремума.
61. Общая задача математического программирования. Функция Лагранжа.
62. Теорема Каруша - Куна - Таккера в форме необходимых условий решения задачи математического программирования.
63. Условие Слейтера.
64. Задача выпуклого программирования.
65. Двойственная функция в задаче математического программирования.
66. Двойственная задача для задачи математического программирования.
67. Сильная двойственность. Зазор двойственности.
68. Локальный анализ чувствительности с помощью множителей Лагранжа.
69. Задача линейного программирования. Задача линейного программирования в стандартной форме.
70. Возможные случаи двойственности в задаче линейного программирования.
71. Симплекс метод.
72. Нахождение первоначальной угловой точки с помощью двухфазного симплекс метода.
73. Сходимость симплекс метода.
74. Показать, что направление антиградиента - направление наискорейшего локального убывания функции.
75. Дифференциальное уравнение градиентного потока.
76. Метод градиентного спуска.
77. Наискорейший спуск.
78. Липшицева парабола для гладкой функции.
79. Размер шага наискорейшего спуска для квадратичной функции.
80. Характер сходимости градиентного спуска к локальному экстремуму для гладких невыпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
81. Характер сходимости градиентного спуска для гладких выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
82. Характер сходимости градиентного спуска для гладких и сильно выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
83. Связь спектра гессиана с константами сильной выпуклости и гладкости функции.
84. Условие Поляка-Лоясиевича (градиентного доминирования) для функций.
85. Сходимость градиентного спуска для сильно выпуклых квадратичных функций. Оптимальные гиперпараметры.
86. Связь PL-функций и сильно выпуклых функций.
87. Привести пример выпуклой, но не сильно выпуклой задачи линейных наименьших квадратов

(возможно, с регуляризацией).
88. Привести пример сильно выпуклой задачи линейных наименьших квадратов (возможно, с регуляризацией).
89. Привести пример выпуклой негладкой задачи линейных наименьших квадратов (возможно, с регуляризацией).
90. Субградиент. Субдифференциал.
91. Субградиентный метод.
92. Характер сходимости субградиентного метода для негладких выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
93. Нижние оценки для гладкой выпуклой оптимизации с помощью методов первого порядка в терминах $\mathcal{O}$ от числа итераций метода.
94. Отличие ускоренной и неускоренной линейной сходимости для методов первого порядка.
95. Метод тяжелого шарика (Поляка).
96. Ускоренный градиентный метод Нестерова для выпуклых гладких функций.
97. Ускоренный градиентный метод Нестерова для сильно выпуклых гладких функций.
98. Проекция.
99. Достаточное условие существования проекции точки на множество.
100. Достаточное условие единственности проекции точки на множество.
101. Метод проекции градиента.
102. Критерий проекции точки на выпуклое множество (Неравенство Бурбаки-Чейни-Гольдштейна).
103. Проекция как нерастягивающий оператор.
104. Метод Франк-Вульфа.
105. Характер сходимости метода проекции градиента для гладких выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
106. Характер сходимости метода проекции градиента для гладких сильно выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
107. Характер сходимости метода Франк-Вульфа для гладких выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
108. Характер сходимости метода Франк-Вульфа для гладких сильно выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
109. A-сопряженность двух векторов. A-ортогональность. Скалярное произведение $\langle \cdot, \cdot \rangle_A$.
110. Процедура ортогонализации Грама-Шмидта.
111. Метод сопряженных направлений.
112. Метод сопряженных градиентов.
113. Зависимость сходимости метода сопряженных градиентнов от спектра матрицы.
114. Характер сходимости метода сопряженных градиентов в терминах $\mathcal{O}$ от числа итераций метода.
115. Метод Поляка-Рибьера.
116. Метод Ньютона.
117. Сходимость метода Ньютона для квадратичной функции.
118. Характер сходимости метода Ньютона для сильно выпуклых гладких функций - куда и как сходится.
119. Демпфированный метод Ньютона.
120. Идея квазиньютоновских методов. Метод SR-1.
121. Нижние оценки для негладкой выпуклой оптимизации с помощью методов первого порядка в терминах $\mathcal{O}$ от числа итераций метода.
122. Проксимальный оператор.
123. Оператор проекции как частный случай проксимального оператора.
124. Характер сходимости проксимального градиентного метода для гладких выпуклых функций f в терминах $\mathcal{O}$ от числа итераций метода.
125. Характер сходимости проксимального градиентного метода для гладких сильно выпуклых

функций f в терминах $\mathcal{O}$ от числа итераций метода.
126. Аналитическое выражение для $\text{prox}_{\lambda\|x\|_1}$.
127. Аналитическое выражение для $\text{prox}_{\frac{\mu}{2}\|x\|_2^2}$.
128. Проксимальный оператор как нерастягивающий оператор.
129. Характер сходимости ускоренного проксимального градиентного метода для гладких выпуклых функций f в терминах $\mathcal{O}$ от числа итераций метода.
130. Метод стохастического градиентного спуска.
131. Идея мини-батча для метода стохастического градиентного спуска. Эпоха.
132. Характер сходимости стохастического градиентного спуска для гладких выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
133. Характер сходимости стохастического градиентного спуска для гладких PL-функций в терминах $\mathcal{O}$ от числа итераций метода.
134. Характер работы стохастического градиентного спуска с постоянным шагом для гладких PL-функций.
135. Основная идея методов уменьшения дисперсии.
136. Метод SVRG.
137. Метод SAG.
138. Метод Adagrad.
139. Метод RMSProp.
140. Метод Adadelta.
141. Метод Adam.
142. Идея проекции функции потерь нейронной сети на прямую, плоскость.
143. Grokking.
144. Double Descent.
145. Взрыв/Затухание градиентов при обучении глубоких нейронных сетей.
146. Идея gradient checkpointing.
147. Идея аккумуляции градиентов.
148. Зачем увеличивать батч при обучении больших нейросетевых моделей. Warmup.
149. Data Parallel обучение на нескольких видеокартах.
150. GPipe Pipeline параллелизм.
151. PipeDream Pipeline параллелизм.
152. Дообучение больших моделей с помощью LoRA адаптеров.
153. Метод двойственного градиентного подъема.
154. Связь константы сильной выпуклости f и гладкости f^*.
155. Идея dual decomposition.
156. Метод двойственного градиентного подъема для линейных ограничений-неравенств.
157. Метод модифицированной функции Лагранжа.
158. Метод ADMM.
159. Формулировка задачи линейных наименьших квадратов с ℓ_1 регуляризацией в форме ADMM.
160. Формулировка задачи поиска точки на пересечении двух выпуклых множеств в форме ADMM.

Теоремы с доказательствами

1. Критерий положительной определенности матрицы через знаки собственных значений матрицы.
2. Автоматическое дифференцирование. Вычислительный граф. Forward/ Backward mode (в этом вопросе нет доказательств, но необходимо подробно описать алгоритмы).
3. Метод дихотомии и золотого сечения для унимодальных функций. Скорость сходимости.
4. Базовые операции, сохраняющие выпуклость множеств: пересечение бесконечного числа множеств, линейная комбинация множеств, образ афинного отображения.
5. Неравенство Йенсена для выпуклой функции и выпуклой комбинации точек.
6. Выпуклость надграфика как критерий выпуклости функции.
7. Дифференциальный критерий сильной выпуклости первого порядка.
8. Дифференциальный критерий сильной выпуклости второго порядка.
9. Необходимые условия безусловного экстремума.
10. Достаточные условия безусловного экстремума.
11. Формулировка симплекс метода для задачи линейного программирования в стандартной форме. Теорема о проверке оптимальности решения.
12. Теорема сходимости градиентного спуска для гладких выпуклых функций.
13. Теорема сходимости градиентного спуска для гладких PL функций.
14. Теорема сходимости градиентного спуска для сильно выпуклых квадратичных функций. Оптимальные гиперпараметры.
15. Теорема сходимости субградиентного метода для выпуклых функций. Сходимость метода для разных стратегий выбора шага: постоянный размер шага $\alpha_k = \alpha$; Обратный квадратный корень $\frac{R}{G\sqrt{k}}$; Обратный $\frac{1}{k}$; Размер шага Поляка: $\alpha_k = \frac{f(x^k)-f^*}{\|g_k\|_2^2}$.
16. Теорема о сходимости метода тяжелого шарика для сильно выпуклой квадратичной задачи.
17. Теорема о сходимости метода проекции градиента для выпуклой гладкой функции.
18. Теорема о сходимости метода проекции градиента для сильно выпуклой гладкой функции.
19. Теорема о сходимости метода Франк-Вульфа для выпуклой гладкой функции.
20. Доказательство сходимости метода сопряженных градиентов и вывод формулы.
21. Теорема сходимости метода Ньютона для сильно выпуклых функций с Липшицевым гессианом.
22. Вывод формул обновления оценок обратного гессиана и гессиана квазиньютоновских методов SR-1, DFP, BFGS.
23. Теорема о сходимости проксимального градиентного метода для выпуклой гладкой функции f.
24. Теорема о сходимости проксимального градиентного метода для сильно выпуклой гладкой функции f.
25. Теорема о сходимости стохастического градиентного спуска в гладком PL-случае.