1. Выберите правильное утверждение.

(a) Централизованная архитектура является масштабируемой, поскольку коммуникации с сервером занимают мало времени.

(b) Коммуникационные протоколы типа AllReduce являются масштабируемыми, поскольку их используют исключительно в датацентрах.

(c) Коммуникационные протоколы типа AllReduce являются масштабируемыми, поскольку каждый рабочий в таких протоколах суммарно передаёт и принимает количество информации соизмеримое с O(1) векторов.

(d) Децентрализованная архитектура плохо масштабируется, поскольку скорость работы зависит от числа обусловленности графа, которое всегда растёт линейно с ростом числа участников n.

Ответ: c

2. Рассмотрим задачу распределённой оптимизации, удовлетворяющую предположениям со слайда 25 первой части лекции 9. Пусть $n = 1000, d = 10^8, L = 10^6, \mu = 10^{-2}, \sigma = 10$ и $\zeta_* = 10^2$. Рассмотрим методы QSGD и DIANA с оператором компрессии $Q(x) = rand_t(x)$, где $t = 10^5$. Оцените число итераций k (в худшем случае), которое необходимо проделать указанным алгоритмам, чтобы гарантировать $\mathbb{E}[\|x^k - x^*\|^2] \leq \varepsilon$ для $\varepsilon = 10^{-3}$ (логарифмами можно пренебречь).

(a) QSGD: порядка 10^{11} итераций; DIANA: порядка 10^9 итераций

(b) QSGD: порядка 10^{11} итераций; DIANA: порядка 10^8 итераций

(c) QSGD: порядка 10^{15} итераций; DIANA: порядка 10^9 итераций

(d) QSGD: порядка 10^9 итераций; DIANA: порядка 10^{11} итераций

$$\omega = \frac{d}{t} - 1 = \frac{10^8}{10^5} - 1 = 999$$

$$QSGD: \quad O\left(\left(1 + \frac{\omega}{n}\right)\frac{L}{\mu}\ln\frac{R_0^2}{\varepsilon} + \frac{(\omega\zeta_*^2 + (\omega+1)\sigma^2)\ln\frac{D_1}{\mu^2\varepsilon}}{n\mu^2\varepsilon}\right) \approx$$

$$\approx O\left(\left(1 + \frac{\omega}{n}\right)\frac{L}{\mu} + \frac{\omega\zeta_*^2 + (\omega+1)\sigma^2}{n\mu^2\varepsilon}\right) = \left(1 + \frac{999}{10^3}\right)\frac{10^6}{10^{-2}} + \frac{999\cdot 10^4 + 10^3\cdot 10^2}{10^3\cdot 10^{-4}\cdot 10^{-3}} \approx 10^{11}$$

$$DIANA: \quad O\left(\left(\omega + \left(1 + \frac{\omega}{n}\right)\frac{L}{\mu}\right)\ln\frac{R_0^2}{\varepsilon} + \frac{(\omega+1)\sigma^2\ln\frac{(1+\omega)\sigma^2}{n\mu^2\varepsilon}}{n\mu^2\varepsilon}\right) \approx$$

$$\approx O\left(\omega + \left(1 + \frac{\omega}{n}\right)\frac{L}{\mu} + \frac{(\omega+1)\sigma^2}{n\mu^2\varepsilon}\right) = 999 + \left(1 + \frac{999}{10^3}\right)\frac{10^6}{10^{-2}} + \frac{10^3\cdot 10^2}{10^3\cdot 10^{-4}\cdot 10^{-3}} \approx 10^9$$

Ответ: a

3. Выберите правильное утверждение.

(a) Методы с локальными шагами не имеют преимуществ перед обычными методами с мини-батчингом в случае гетерогенных функций; Local-SGD хорошо работает для постановок задач с персонализацией

(b) Методы с локальными шагами всегда работают хуже, чем методы с мини-батчингом, поскольку локальные шаги создают "дрифт"рабочих (каждый рабочий стремится к своему оптимуму)

(c) Методы с локальными шагами работают лучше, чем обычные методы с мини-батчингом, поскольку меньше коммуницируют с сервером

(d) Локальные шаги работают только в применении к стандартному SGD

Ответ: a

4. Какая из приведенных ниже матриц не может являться коммуникационной матрицей (mixing matrix) некоторого графа?

(a) $\begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix}$

(b) $\begin{pmatrix} 2/3 & 1/3 & 1/3 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 1/3 & 2/3 \\ 0 & 0 & 2/3 & 1/3 \end{pmatrix}$

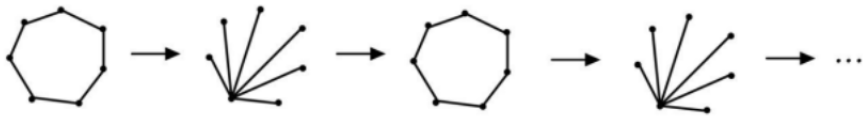
(c) $\begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$

Условию $\mathbf{M1} = \mathbf{1}$ не удовлетворяет только матрица b

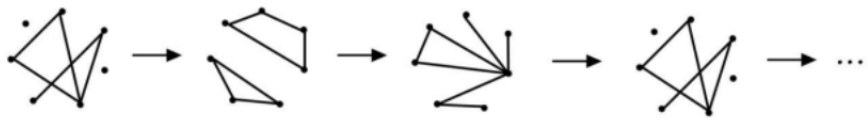
Ответ: b

5. В силу технических неполадок в коммуникационной сети иногда возникают и пропадают связи между узлами. Каждый узел хранит функцию f_i и локальный вектор параметров x_i, требуется минимизировать сумму $\sum_{i=1}^{m} f_i(x_i)$ при сохранении (приблизительного) консенсуса $x_1 = \cdots = x_m$. В каком из случаев заведомо не получится добиться консенсуса между узлами?

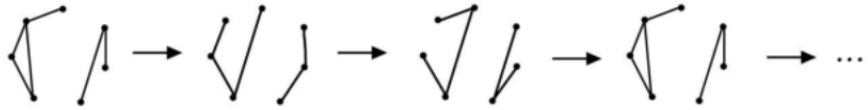
(a)



(b)



(c)



Только в случае *c* объединение подряд идущих графов является несвязным графом (две компоненты)
Ответ: *c*

Ответы:
1. c
2. a
3. a
4. b
5. c