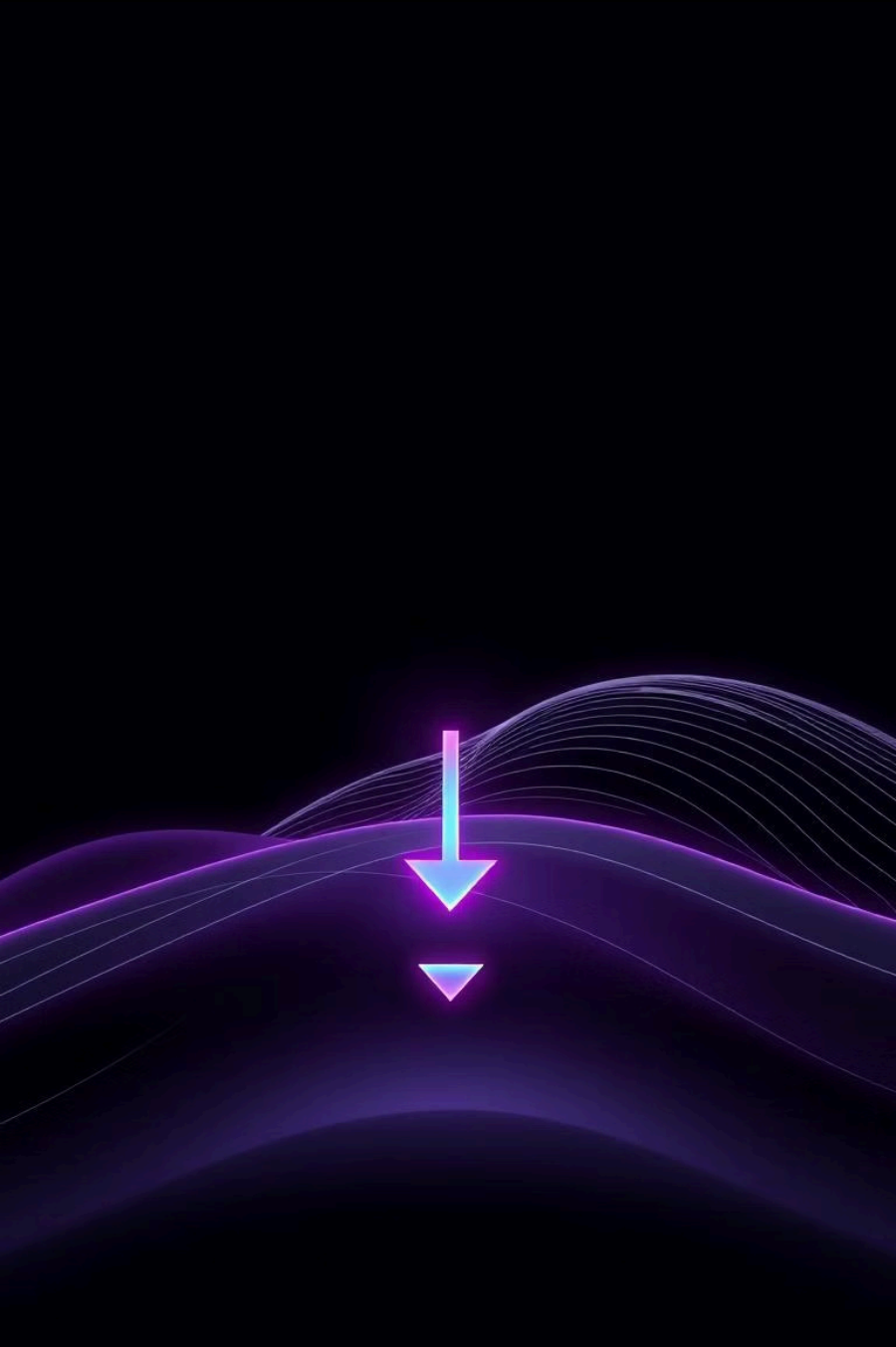




Дәріс 6: Градиенттік Төмендеу және Оңтайландыру

Оқытушы: Сарсембаева Т.С.

Негізгі Сұрақтар

1

Градиенттік төмендеу

Машиналық оқытудағы негізгі мақсаты қандай?

2

Градиент бағыты

Алгоритм неліктен градиентке қарсы бағытта қозғалады?

3

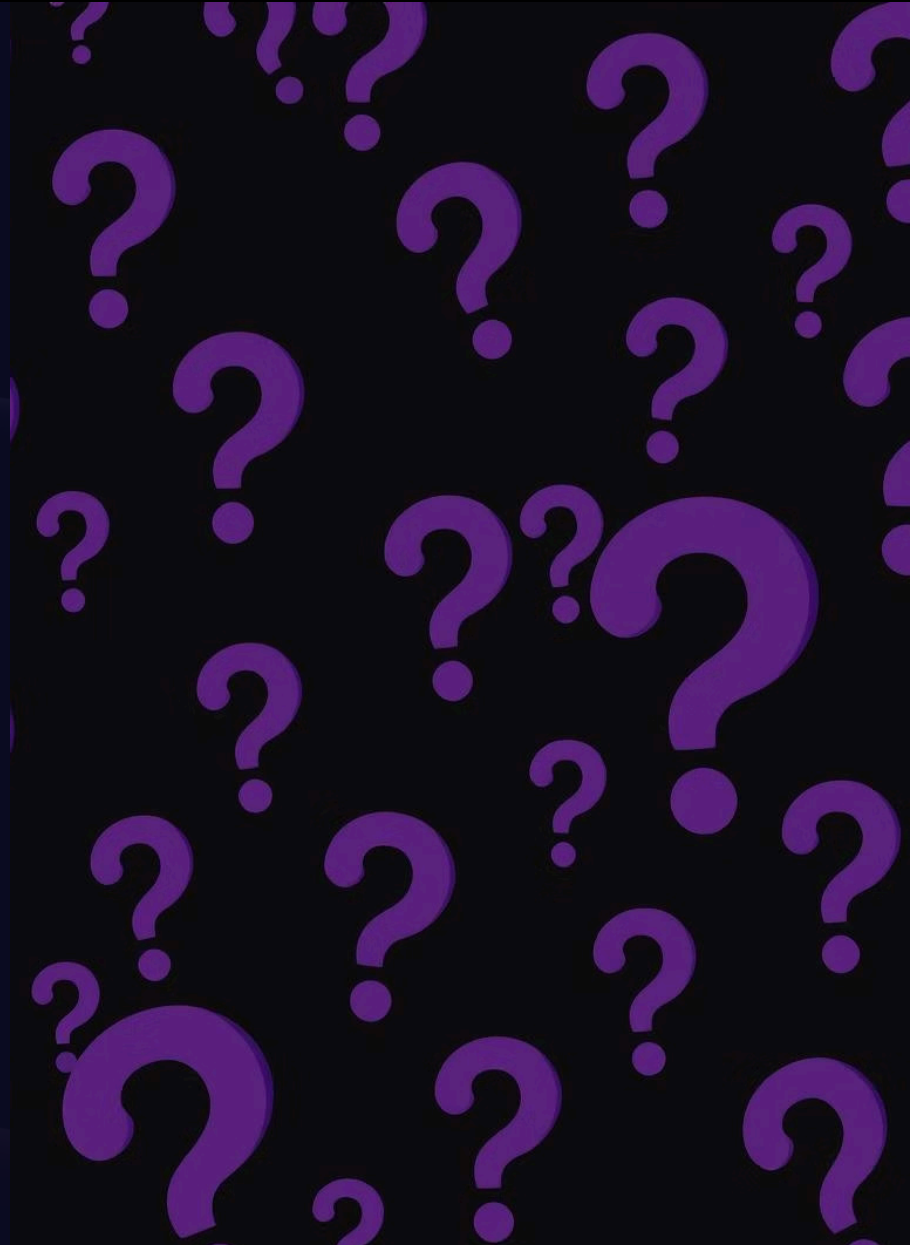
Оқыту жылдамдығы

Тығыз үлкен немесе тым кіші болса не болады?

4

GD түрлері

Batch, Stochastic (SGD) және Mini-Batch GD арасындағы айырмашылық неде?



Градиенттік Төмендеу: Негізгі Түсініктер

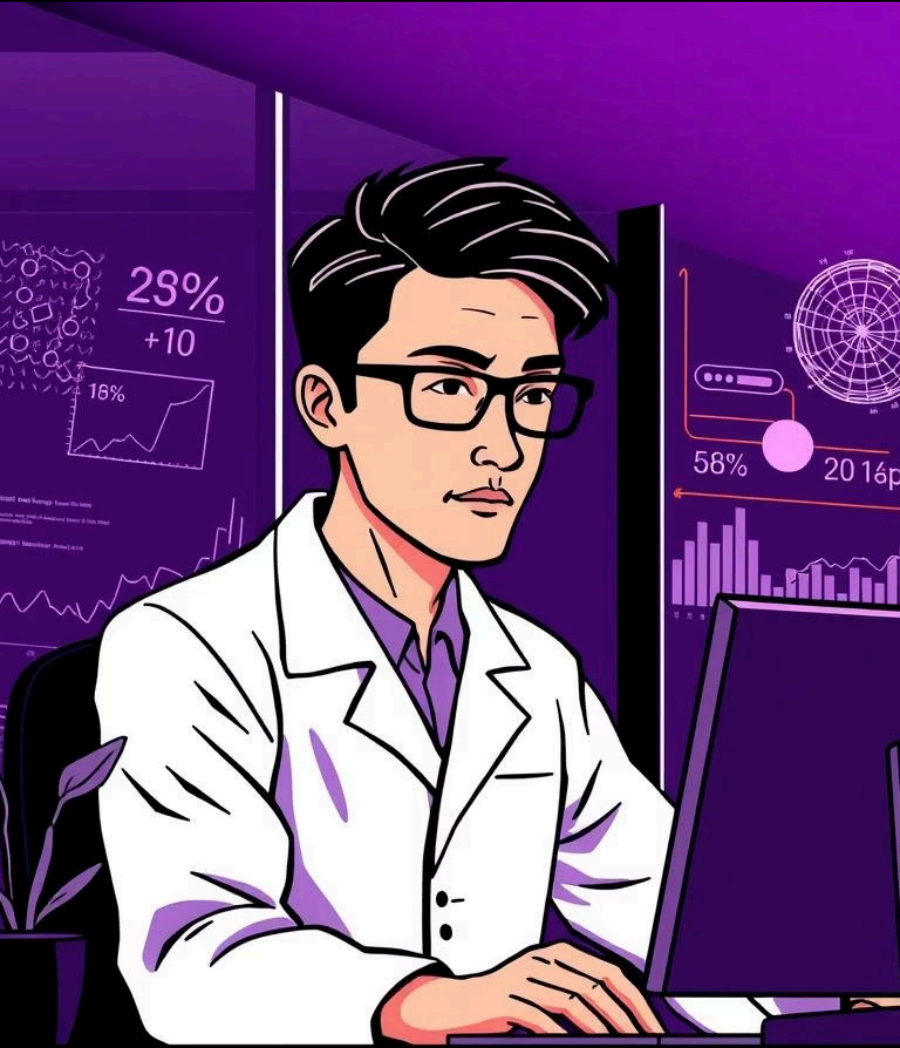
Машиналық оқыту алгоритмдері оңтайландыруға негізделген. Градиенттік төмендеу – сызықтық регрессиядан терең нейрондық желілерге дейінгі көптеген модельдердің негізгі құралы. Бұл қателік (шығын) функциясының мәнін азайту үшін қолданылатын итерациялық оңтайландыру алгоритмі.

Оның негізгі идеясы – алгоритм функцияның ең жылдам өсу бағытын (градиент) табады. Функцияны азайту үшін ол сол градиентке қарама-қарсы бағытта жылжиды.

Оқыту Жылдамдығы (Learning Rate)

- Қадам мөлшерін анықтайды, алгоритмнің тұрақтылығы мен жинақылық жылдамдығына тікелей әсер етеді.
- **Тығыз үлкен болса:** алгоритм минимумнан "секіріп" өтіп, тұрақсыз болуы немесе шығынның өсуі мүмкін.
- **Тығыз кіші болса:** алгоритм өте баяу жылжиды, бұл есептеу уақытын көп қажет етеді.





Кіріспе: Деректер Ғылымындағы Оңтайландыру

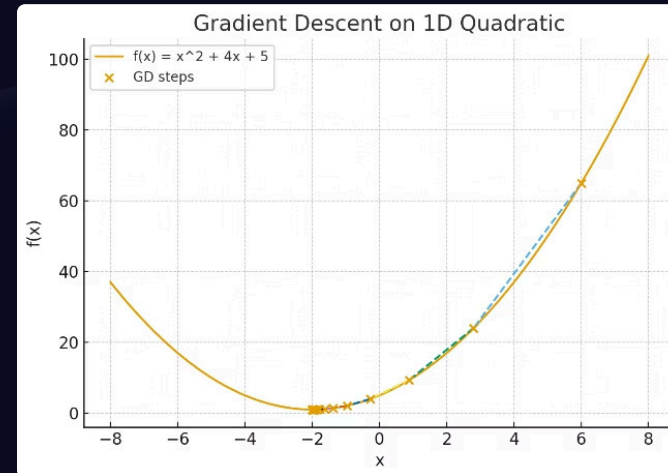
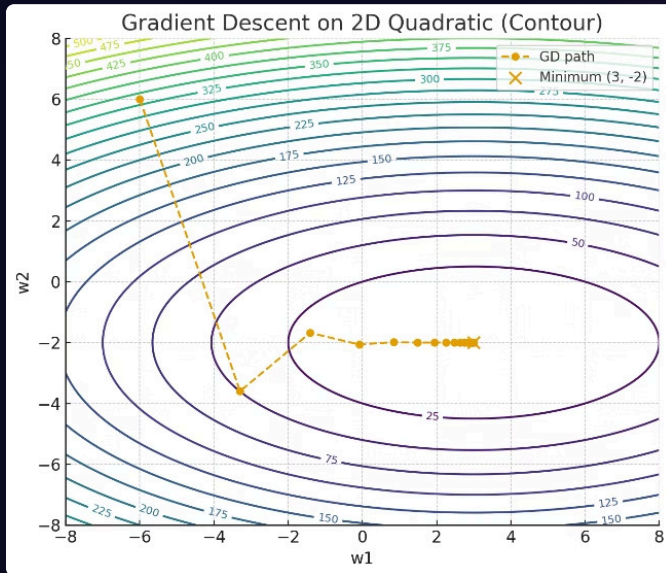
Деректер ғалымының негізгі жұмысы — қол жетімді деректердегі үлгілерді сипаттайтын және жаңа байқаулар үшін ақылға қонымды болжамдар бере алатын идеалды модельді табу.

Көптеген машиналық оқыту алгоритмдері оңтайландыруға негізделгендіктен, функцияларды азайту және көбейту жиі кездеседі. Мұны жасау үшін ең негізгі және жалпы құралдардың бірі — градиентті төмендеу.

Бұл тәсіл бүгінгі таңда қолданатын көптеген модельдердің негізі болып табылады (сызықтық регрессиядан күрделі терең нейрондық желілерге дейін), және оны түсіну бізге деректер ғылымы әлемінде модельдерді саналы түрде құруға, оқытуға және жетілдіруге мүмкіндік береді.

Градиенттік Төмендеудің Негіздері

Градиентті төмендету - қатені немесе шығын функциясының мәнін азайту үшін қолданылатын итерациялық оңтайландыру алгоритмі. Математикалық деңгейде бұл көп өлшемді функцияның оңтайландыру тапсырмасы болып табылады.



Градиенттердің түсу тұжырымдамасы көп айнымалы функцияны азайту мақсатында оның теріс градиент бағытына қарай жүру болып табылады. Градиент - бұл функцияның максималды өсу бағыты, сол себепті оның қарама-қарсы бағыты максималды азаю бағыты болып табылады.

Жаңартылған параметрлер төмендегідей беріледі:

$$\theta_{new} = \theta_{old} - \eta \cdot \nabla L(\theta)$$

- θ_{new} — модель параметрі (мысалы, салмақтар w және қиян b)
- η — оқыту жылдамдығы, ол қадам ұзындығын білдіреді.
- $\nabla L(\theta)$ — жоғалту функциясындағы градиент.

Оқыту Жылдамдығының (Learning Rate) Маңызы

Оқу жылдамдығы (Learning rate) - градиенттік төмендеу алгоритмі арқылы параметрлерді жаңартуда қолданылатын қадам мөлшері.

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t)$$

Яғни, әр қадам сайын параметр мәні кішігірім вектор $\eta \nabla L(\theta_t)$ арқылы жаңартылады.

Оқу жылдамдығының әсері:

Тұрақтылық пен Жинақылық

Алгоритмнің тұрақтылығы мен жинақылық жылдамдығы тікелей η -ға байланысты.

Шығын Беті

Шығын беті әдетте үйлесімсіз: кейбір жерлерде өте қисық, ал кейбір жерлерде өте тік болады.

Тығыз Үлкен η

Алгоритм минималды «асырып» жібереді; итерация барысында loss өсіп немесе тұрақсыз шуы пайда болады.

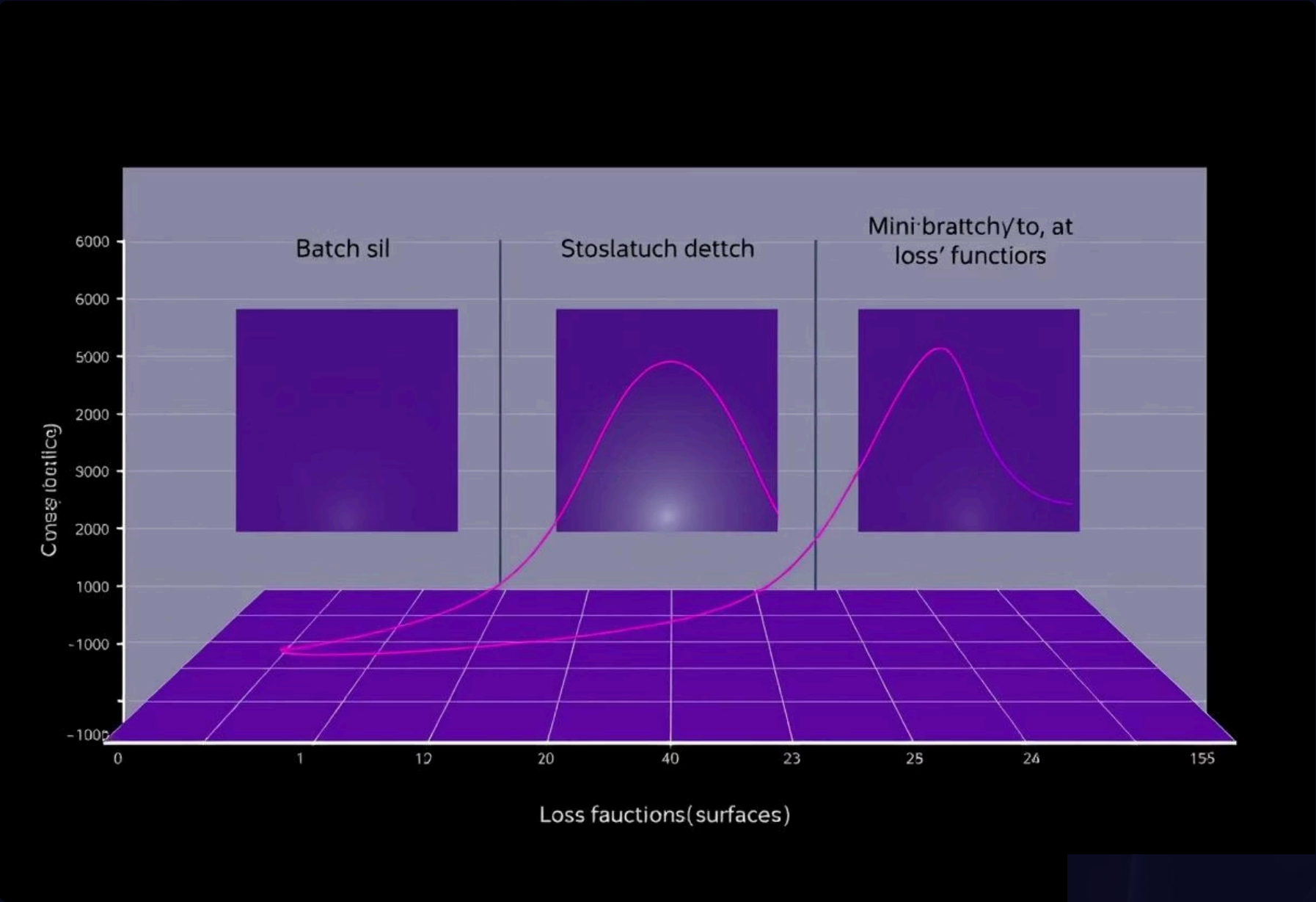
Тығыз Кіші η

Әр градиенттік төмендеу қадамы өте кішкентай, өте баяу алға жылжиды; есептеу ресурстары мен уақыттық мүмкіндігінен іс жүзінде асып кетуі мүмкін.

Градиенттік Төмендеудің Түрлері

Градиентті төмендеудің үш жалпы түрі бар: Batch Gradient Descent, Stochastic Gradient Descent және Mini-Batch Gradient Descent.

Batch Gradient Descent	Толық деректер жиынтығын пайдаланады. Әрбір итерацияда қате мәні модель бойынша қолданылады.	Дәл нәтиже береді, тегіс түсу траекториясы бар. Жаһандық минимумға жету ықтималдығы жоғары.	Үлкен деректер жиынтығы үшін баяу, көп RAM қажет.	Regression, Neural Networks (шағын dataset)
Stochastic Gradient Descent (SGD)	Әрбір итерацияда градиент тек бір үлгі үшін есептеледі.	Нақты уақыт (онлайн оқыту) есептеулеріне қатысты өте жылдам. Жергілікті минимумнан шығуға жақсы мүмкіндік береді.	Траектория тұрақсыз «тербеледі». Тұрақсыз және кейде баяу біріктіріледі.	Online Learning, Deep Learning, Recommendation Systems
Mini-Batch Gradient Descent	Деректер жиынтығын көптеген кіші топтарға бөліп, әрбірі үшін градиентті есептейді.	Тепе-теңдік сақтайды: жылдам және тұрақты. GPU-да параллель есептеулер үшін ыңғайлы. SGD-дегі "шуды" азайтады.	Топ мөлшерін икемдеу өте маңызды (әдетте 32-512).	CNN, NLP, Deep Learning (TensorFlow, PyTorch)



Әдебиеттер Тізімі

- Kordowich, G., & Jaeger, J. (2023). A Physics Informed Machine Learning Method for Power System Model Parameter Optimization.
<https://arxiv.org/pdf/2309.16579>
- «Лекция 9. Градиентный спуск и его модификации для решения задачи оптимизации». (2022, 18 февраля). Курс: «Математика для анализа данных». Лекторы: И. Оселедец, А. Катруца. Teach-In AI.
<https://teach-in.ru/lecture/2022-02-18-Matematiki-1>
- «Gradient Descent: Основы Data Science». NeuroHive.
<https://neurohive.io/ru/osnovy-data-science/gradient-descent/>
- Ruder, S. (2016). An overview of gradient descent optimization algorithms.
<https://arxiv.org/abs/1609.04747>
- Jin, C., Netrapalli, P., Ge, R., Kakade, S. M., & Jordan, M. I. (2019). On Nonconvex Optimization for Machine Learning: Gradients, Stochasticity, and Saddle Points. <https://arxiv.org/abs/1902.04811>

