

## Notebook 1

در این تمرین عملی به بررسی مبانی بهینه‌سازی مبتنی بر گرادین در یادگیری عمیق می‌پردازیم و سه نوع optimizer (SGD, SGD+momentum, SGD+nesterov) را پیاده‌سازی و مقایسه می‌کنیم. دانشجویان باید بخش‌های مشخص‌شده در کد با کامنت TODO را تکمیل کنند. در این تمرین یک مدل CNN بر روی دیتاست CIFAR-10 با batch-size ها و optimizer های مختلف آموزش می‌دهیم. دانشجویان سه آزمایش جامع انجام خواهند داد:

1. بررسی تأثیر اندازه batch بر همگرایی و تعمیم‌پذیری
  2. مقایسه انواع momentum برای درک چگونگی تسریع آموزش
  3. تحلیل حساسیت نرخ یادگیری در optimizer های مختلف
- پس از تکمیل کد و اجرای آن، یک گزارش بنویسید که در آن تمامی نمودارها و نتایج را بررسی و تحلیل کنید که برای هر experiment باید شامل موارد زیر باشد:

- Experiment 1: در این قسمت باید به تحلیل تأثیر اندازه batch های مختلف شامب سرعت همگرایی، تحلیل کارایی و پایداری و مقایسه full-batch و mini-batch در GD پرداخت.
- Experiment 2: در این قسمت انواع momentum و تأثیر آن بر همگرایی، رفتار گرادین، دقت در هر کلاس و مقایسه momentum و Nesterov باید بررسی شود.
- Experiment 3: در این قسمت این سه optimizer را بر اساس میزان حساسیت به نرخ یادگیری باید بررسی و به صورت کامل تحلیل شود.

## Notebook 2

در این تمرین عملی به بررسی optimizer های با نرخ یادگیری تطبیق

(Adaptive Learning Rate Optimizers) در یادگیری عمیق می‌پردازیم و سه نوع Optimizer (Adagrad, RMSProp, Adam) را پیاده‌سازی و مقایسه می‌کنیم. دانشجویان باید بخش‌های مشخص‌شده در کد با کامنت TODO را تکمیل کنند. یک مدل MLP بر روی دیتاست Fashion-MNIST با Optimizer های تطبیقی مختلف آموزش داده می‌شود. دانشجویان سه آزمایش انجام می‌دهند:

1. مقایسه جامع optimizer های تطبیقی (Adagrad vs RMSProp vs Adam)
2. تحلیل استحکام در برابر نرخ یادگیری (LR Robustness) - بررسی حساسیت هر optimizer به انتخاب نرخ یادگیری

### 3. تحلیل حساسیت Hyperparameter های Adam (beta1, beta2, epsilon)

پس از تکمیل کد و اجرای آن، یک گزارش بنویسید که در آن تمامی نمودارها و نتایج را بررسی و تحلیل کنید که برای هر experiment باید شامل موارد زیر باشد:

- Experiment 1 (مقایسه Optimizer های تطبیقی): در این قسمت باید به تحلیل سرعت همگرایی، دقت نهایی و پایداری، رفتار نرخ یادگیری موثر، تحلیل گرادیانها و بزرگی بهروزرسانی پارامترها، دقت per-class و مقایسه مزایا و معایب Adagrad، RMSProp و Adam پرداخته شود.
- Experiment 2 (استحکام در برابر نرخ یادگیری): در این قسمت حساسیت هر optimizer به learning rate، محدوده قابل قبول LR، رفتار با LR بالا/پایین و توصیه عملی برای انتخاب optimizer باید بررسی شود.
- Experiment 3 (حساسیت Hyperparameter های Adam): در این قسمت تاثیر Beta1 بر سرعت و پایداری، نقش Beta2 در تطبیق نرخ یادگیری، اهمیت Epsilon و trade-off بین hyperparameter ها باید به صورت کامل تحلیل شود.