

تنظیم نرخ یادگیری Learning Rate Scheduling

چند روش رایج برای تنظیم نرخ یادگیری :

1-کاهش چندجمله‌ای

Polynomial Decay

در این روش، نرخ یادگیری به صورت چندجمله‌ای کاهش می‌یابد. این روش انعطاف پذیری بیشتری نسبت به کاهش نمایی دارد.

فرمول :

$$lr = lr0 * (1 - t/T)^P$$

T : تعداد کل اپوک ها

P : توان که معمولا عددی ثابت است

مزایا :

انعطاف پذیری: به کمک پارامتر توان، می‌توان شیب کاهش نرخ یادگیری را کنترل کرد.

بهبود همگرایی: به مدل اجازه می‌دهد تا در مراحل پایانی آموزش با نرخ یادگیری کوچک‌تری بهینه‌سازی شود.

2-کاهش کسینوسی

Cosine Annealing

در این روش، نرخ یادگیری به صورت نوسانی با کاهش تدریجی کاهش می‌یابد.

فرمول :

$$lr = lr0 / 2 * (1 + \cos(\pi * t / T))$$

T : تعداد کل اپوک ها

مزایا :

پایداری: با کاهش تدریجی نرخ یادگیری به صورت نوسانی، مدل می‌تواند به همگرایی بهتری برسد.

پیشگیری از گیرافتادن در مینیمم‌های محلی: با تغییر نرخ یادگیری به صورت نوسانی، مدل می‌تواند از گیرافتادن در مینیمم‌های محلی جلوگیری کند.