



مسئله ۱. Backpropagation

در یک شبکه عصبی سنتی، ورودی لایه $(l+1)$ به صورت $x_{l+1} = F(x_l)$ بدست می آید که F به صورت خلاصه بیانگر ضرب وزن ها و اعمال تابع فعال سازی است. حال فرض کنید ورودی لایه $(l+1)$ را بدین صورت محاسبه کنیم:

$$x_{l+1} = F(x_l) + x_l$$

الف

اگر شبکه در مجموع L لایه داشته باشد، رابطه x_L بر حسب x_l را بدست آورید.

ب

اگر تابع هزینه این شبکه عصبی را E بنامیم، عبارت Backpropagation روبرو $\frac{\partial E}{\partial x_l}$ را محاسبه کنید.

پ

این معماری مشابه معماری ResNet است که برای آموزش شبکه های عصبی بسیار عمیق پیشنهاد شده است. با توجه به رابطه ای که در قسمت قبل بدست آوردید، توضیح دهید تفاوت عبارت backpropagation در این معماری در برابر معماری سنتی چیست که امکان عمیق تر کردن شبکه های عصبی را فراهم آورده است؟

مسئله ۲. بهینه سازی در شبکه های عصبی

الف

در ابعاد بالا، احتمال برخورد با نقاط زینی^۱ بسیار بیشتر از احتمال برخورد با نقاط اکسترمم محلی است. توضیح دهید روش های مبتنی بر momentum چه برتری نسبت به روش SGD

^۱saddle point

در برخورد با این نقاط دارند؟

ب

مزیت روش‌های Adaptive مانند Adam و Adagrad نسبت به روش‌های momentum چیست؟

پ

علت استفاده از Bias Correction در روش Adam چیست؟

موفق باشید :)