

بسمه تعالی

۱. زمانی که از روش Stochastic Gradient Descent استفاده می‌کنیم. ممکن است شاهد افزایش مقدار تابع هزینه پس از بروزرسانی وزن‌ها باشیم. چرا این اتفاق ممکن است رخ دهد؟ (دو دلیل را ذکر کنید.)

۲. برای گیت‌های منطقی AND و OR با سه ورودی، شبکه‌ی پرسپترون طراحی کنید.

۳. در کلاس دیدیم با نگاه احتمالاتی به مسئله‌ی رگرشن و با فرض $p(y|f(x; W)) \sim \mathcal{N}(f(x; W), \sigma^2)$ می‌توان بیشینه کردن درستنمایی را معادل با تابع هزینه‌ی میانگین مربعات خطا در نظر گرفت. حال فرض کنید $p(y|f(x; W)) \sim \mathcal{L}(f(x; W), b)$ آمده باشد. رابطه‌ی توزیع لاپلاس به صورت زیر است:

$$\mathcal{L}(\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right), \quad b > 0 \quad (1)$$

در این صورت تابع هزینه‌ای که از بیشینه کردن درستنمایی بدست می‌آید به چه صورت است؟