به نام خدا

درس یادگیری ماشین
نیم‌سال دوم ۰۳-۰۴
استاد: ربیعی- دهقانیان

دانشکده مهندسی کامپیوتر

آزمونک اول - ۱۴۰۳/۱۲/۱۹ زمان آزمون: ۳۰ دقیقه

---

۱. (۴۰ نمره) به جای استفاده از خطای میانگین مربعات ،(MSE) فرض کنید که یک مدل رگرسیون خطی را با کمینه کردن خطای میانگین مکعبی آموزش دهید:

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\theta^T x^{(i)} - y^{(i)}\right)^3$$

(الف) گرادیان خطای میانگین مکعبی نسبت به بردار پارامتر θ چیست؟

(ب) آیا این تابع هزینه نسبت به خطای میانگین مربعات حساسیت بیشتری به نقاط پرت دارد ؟

(ج) در ۱ تا ۲ جمله مختصر، به طور خلاصه مشکل اصلی استفاده از این تابع هدف برای رگرسیون را توضیح دهید.

۲. (۳۰ نمره) در این مسئله، تأثیر افزایش پیچیدگی نمایش ویژگی‌ها را در حالی که اندازه‌ی مجموعه‌ی آموزشی ثابت است، بررسی می‌کنیم.

یک مسئله‌ی دسته‌بندی روی خط حقیقی R را در نظر بگیرید که در آن داده‌ها از توزیع D گرفته شده و بر اساس تابع هدف $c^* : R \to \{\pm 1\}$ دسته‌بندی می‌شوند. فرض کنید مجموعه‌ی آموزشی S شامل n نمونه‌ی تصادفی و هم‌توزیع از D در اختیار داریم.

برای هر مقدار d، یک نمایش جدید از داده‌ها را با استفاده از نگاشت ویژگی ϕ_d به‌صورت زیر تعریف می‌کنیم:

$$\phi_d(x) = (1, x, x^2, \ldots, x^d)$$

این نگاشت هر نقطه روی خط حقیقی را به فضایی $(d+1)$- بعدی منتقل می‌کند. فرض کنید که پس از اعمال این نگاشت، یک روش دسته‌بندی ساده برای تفکیک داده‌ها به کار گرفته می‌شود (مثلاً جداسازی بر اساس یک قاعده‌ی ساده در فضای جدید).

(الف) آیا ممکن است که با افزایش درجه‌ی d در نگاشت ویژگی، **خطای دسته‌بندی روی داده‌های آموزشی** افزایش یابد؟

(ب) توضیح دهید که چرا **خطای دسته‌بندی روی داده‌های جدید** ابتدا کاهش می‌یابد اما پس از مدتی با افزایش d دوباره رشد می‌کند.

۳. (۳۰ نمره) یک متغیر تصادفی مانند X تعداد نمونه به صورت i.i.d. در دست هستند. در مورد خانواده توزیع X چیزی نمی‌دانیم، اما هدف ما تنها تخمین مقدار احتمال $P(X \geq 0)$ است و به دنبال تخمین پارامترهای بیشتری از توزیع نیستیم.

(الف) با استفاده از نمونه‌ها، تخمین‌گر بیشینه درست‌نمایی را برای مقدار $P(X \geq 0)$ بدست آورید.

(ب) حال فرض کنید که در مورد احتمال خواسته شده اطلاعاتی موجود است که به صورت توزیع پیشین $f(p)$ مدل‌سازی شده است. فرض کنید این توزیع پیشین یک توزیع بتا به فرم:

$$f(p) = kp^{k-1}, \quad 0 \leq p \leq 1$$

با k عددی بزرگتر از ۱ است. در این حالت، تخمین‌گر بیشینه توزیع پسین (MAP) برای مقدار نامعلوم $P(X \geq 0)$ چیست؟

(ج) نشان دهید که تخمین‌گر MAP در این حالت مشابه با تخمین‌گر ML بوده با این تفاوت که انگار تعداد $k-1$ عدد مثبت به n نمونه اولیه اضافه شده‌اند.