

سوال ۱ (۱۰ نمره): تخمین MAP و ML

۱.۱. (۵ نمره) فرض کنید توزیع نمونه‌های $\{x_1, \dots, x_n\}$ گاوسین با ماتریس کوواریانس شناخته شده Σ و میانگین ناشناخته μ باشد. چنانچه $\mu \sim \mathcal{N}(\mu_0, \Sigma_0)$ باشد، $\hat{\mu}_{MAP}$ را پیدا کنید. آیا اگر روی فضای نمونه‌ها تبدیل $x' = A^T x$ اعمال شود $(\det(A) \neq 0)$ ، $\hat{\mu}_{MAP}$ تخمین مناسب برای میانگین داده‌های تبدیل‌یافته $\{x'_1, \dots, x'_n\}$ نیز هست (توضیح دهید)؟

۲.۱. (۵ نمره) فرض کنید متغیر تصادفی m مشخص کننده تعداد رخداد $x = 1$ از بین N تکرار (برای متغیر تصادفی دودویی x با توزیع برنولی که میانگین μ دارد) باشد که از توزیع دو جمله‌ای $Bin(m|N, \mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m}$ پیروی می‌کند و میانگین آن یعنی μ توزیع بتا $Beta(\mu|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1} (1-\mu)^{b-1}$ دارد $\Gamma(x)$ تابع گاما $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$ است. اگر تعداد رخدادهای $x = 1$ برابر m و تعداد رخداد $x = 0$ برابر l باشد $(l = N - m)$ ، نشان دهید مقدار میانگین توزیع پسین (posterior) بین میانگین پیشین (prior) و تخمین $\hat{\mu}_{ML}$ واقع می‌شود. نتیجه‌ی حاصل را توصیف نمایید.

(برای مطالعه توزیع بتا می‌توانید به بخش 2.1.1 کتاب Bishop مراجعه کنید.)

سوال ۲ (۳۰ نمره): مدل‌سازی غیرپارامتری

۱.۲. تابع چگالی یکنواخت $U(0.5, 2.5)$ را در نظر بگیرید:

a. (۷ نمره) نتایج تخمین k -نزدیک‌ترین همسایه را برای $k = 4, 16, 64$ در دو حالتی که $N = 500$ و $N = 2500$ داده تصادفی مطابق توزیع بالا تولید شده باشد، رسم نمایید.

b. (۸ نمره) نتایج تخمین پنجره Parzen را در حالتی که تابع پنجره گاوسی $\mathcal{N}(0, 1)$ باشد را برای پهناهای پنجره $h = 0.01$ ، $h = 0.05$ و $h = 0.2$ و در هر حالت برای تعداد داده $N = 32$ ، $N = 256$ و $N = 2500$ رسم نمایید.

۲.۲. (۷ نمره) یک دسته‌بند طراحی کنید که با گرفتن مجموعه نمونه‌های آموزش از تصمیم بیز (با فرض احتمال پیشین برابر دسته‌ها) برای دسته‌بندی استفاده نماید و چگالی شرطی دسته‌ها را با استفاده از روش پنجره Parzen با پنجره گاوسی $(\sigma = 0.01, \sigma = 0.1, \sigma = 1, \sigma = 10)$ تخمین بزند. در صد خطای این دسته‌بند را روی نمونه‌های آموزش و آزمون موجود در فایل "iris.mat" به ازای پهناهای پنجره‌ی مذکور گزارش کنید.

۳.۲. (۸ نمره) دسته‌بند k نزدیک‌ترین همسایه را پیاده‌سازی کنید و خطای آن را به ازای $k = 1$ ، $k = 5$ و $k = 15$ روی نمونه‌های آموزش و آزمون موجود در فایل "iris.mat" مشخص کنید. در صورتی که از تکنیک 10-fold CV برای پیدا کردن پارامتر k استفاده کنید، مقدار خطای حاصل روی نمونه‌های آزمون و همچنین مقدار k انتخاب شده را تعیین نمایید.

سوال ۳ (۱۵ نمره): خطای دسته‌بند نزدیک‌ترین همسایه

۱.۳. در این قسمت حالاتی که نرخ خطای دسته‌بند نزدیک‌ترین همسایه ($N \rightarrow \infty$) با خطای بیز برابر می‌شود را مورد بررسی قرار می‌دهیم:

a. (۳ نمره) نشان دهید که نرخ خطای دسته‌بند نزدیک‌ترین همسایه P ($N \rightarrow \infty$) با نرخ خطای بیز P^* در دو حالت $P^* = 0$ و $P^* = (c - 1)/c$ برابر خواهد شد.

b. (۶ نمره) برای مسأله‌ی دسته‌بندی به c دسته با احتمال پیشین یکسان و چگالی‌های احتمال شرطی دسته‌ها به صورت زیر:

$$p(x|\omega_i) = \begin{cases} 1 & 0 \leq x \leq cr/(c-1) \\ 1 & i \leq x \leq i+1 - cr/(c-1) \\ 0 & \text{otherwise} \end{cases}$$

نیز نشان دهید که $P = P^* = r$ است.

c. (۱ نمره) نتیجه به دست آمده در قسمت b را به صورت ترکیب دو حالت موجود در قسمت a توصیف کنید.

۲.۳. (۵ نمره) یک مسأله‌ی دسته‌بندی با دو دسته در فضای ویژگی دویعدی در نظر بگیرید. فرض کنید نمونه‌های هر یک از دو دسته داخل دایره‌ای به شعاع r با توزیع یکنواخت پخش شده‌اند و فاصله‌ی مراکز این دو دایره بیش‌تر از $4r$ است. چنانچه N نمونه‌ی آموزش در اختیار داشته باشیم، نشان دهید که احتمال خطای دسته‌بند NN همیشه کمتر از خطای kNN ($k \geq 3$) است.

سوال ۴: (۲۵ نمره) توابع تفکیک خطی

۱.۴. در مسأله‌ی دسته‌بندی چنددسته‌ای (multi-category)، یک مجموعه نمونه "خطی جدایی‌پذیر" (linearly separable) گفته می‌شود اگر یک ماشین خطی (linear machine) وجود داشته باشد که همه‌ی نمونه‌های این مجموعه را درست دسته‌بندی کند. در حالتی که نمونه‌های هر دسته قابل جدا شدن از نمونه‌های بقیه‌ی دسته‌ها توسط یک ابر صفحه باشند نمونه‌ها "کاملاً خطی جدایی‌پذیر" (totally linearly separable) گفته می‌شوند. همچنین یک مجموعه نمونه "دوبه‌دو خطی جدایی‌پذیر" گفته می‌شود اگر نمونه‌های هر زوج دسته توسط یک ابر صفحه قابل جدا شدن باشند.

a. (۴ نمره) نشان دهید نمونه‌های کاملاً جدایی‌پذیر خطی، جداپذیر خطی نیز هستند، اما بالعکس آن لزوماً درست نیست.

b. (۶ نمره) نشان دهید نمونه‌های دوبه‌دو خطی جدایی‌پذیر، لزوماً خطی جدایی‌پذیر نیستند. آیا بالعکس نمونه‌های خطی جدایی‌پذیر لزوماً دوبه‌دو خطی جدایی‌پذیر نیز هستند؟

۲.۴. (۴ نمره) آیا روش حداقل مجموع مجذورات خطا (SSE) همیشه برای یک مسأله دو دسته‌ای که نمونه‌ها خطی جدایی‌پذیرند، قادر است مرز خطی که نمونه‌ها را جدا می‌کند، پیدا کند. در صورت درستی اثبات کنید و در صورت نادرست بودن مثال نقض بیاورید.

۳.۴. (۵ نمره) دسته‌بند پرسپترون (Perceptron) که به روزرسانی وزن را به صورت single-sample و با $\eta = 1$ انجام می‌دهد را در نظر بگیرید. چنانچه بخواهیم مرز دو دسته‌ای که نمونه‌های آن‌ها با درصد خطای کم اما ناصفر قابل جداسازی با ابر صفحه هستند را پیدا کنیم، آیا فقط با تغییر دادن شرط پایانی این روش می‌توانیم به نتیجه مطلوب برسیم؟ آیا تنظیم η با دور تکرار به روزرسانی می‌تواند به نتیجه‌ی بهتری منجر شود (توضیح)؟

۴.۴. (۶ نمره) برای SVM حاشیه نرم (soft-margin) نشان دهید که نمونه‌های با $0 < \alpha_i < C$ دقیقاً روی ابرصفحه‌های حاشیه واقع می‌شوند و نمونه‌های با $\alpha_i = C$ روی حاشیه قرار می‌گیرند یا از آن رد می‌شوند.

سوال ۵ (۲۵ نمره): پیاده‌سازی دسته‌بندی خطی

۱.۵ (۶ نمره) پرسپترون تغییر یافته‌ی پیشنهادی‌تان را که نیاز نداشته باشد دو دسته‌ی ورودی کاملاً خطی جدایی‌پذیر باشند، را پیاده‌سازی کنید. به‌روزرسانی وزن را به صورت single-sample انجام دهید و مقادیر درایه‌های بردار وزن اولیه را صفر در نظر بگیرید.

۲.۵ (۵ نمره) چنانچه از کمینه کردن مجموع مجذورات خطا (SSE) برای پیدا کردن دسته‌بندی خطی استفاده نمایید، بردار وزن به صورت $\hat{\mathbf{w}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ در می‌آید و به دسته‌بندی حاصل شبه‌وارون (pseudoinverse) نیز گفته می‌شود. یک تابع MATLAB برای این منظور پیاده‌سازی کنید که با گرفتن نمونه‌های آموزش، بردار وزن حاصل را برگرداند. نکات لازم در پیاده‌سازی باید رعایت شود که در عمل استفاده از این تابع با مشکل محاسبات عددی مواجه نشود.

۳.۵ (۵ نمره) فایل‌های شامل پیاده‌سازی آموزش SVM حاشیه نرم (soft-margin) و به‌کارگیری دسته‌بندی SVM آموزش دیده برای پیدا کردن دسته یک نمونه جدید را تکمیل نمایید.

۴.۵ (۴ نمره) درصد خطای دسته‌بندی پرسپترون پیشنهادی (۱.۷)، شبه‌وارون و SVM ($C = 0.5$) را روی مجموعه داده‌های آموزش و آزمون موجود در فایل "wdbc.mat" مشخص کنید و نتایج را مقایسه نمایید.

۵.۵ چنانچه برای تنظیم پارامتر C در روش SVM از 10-fold cross-validation استفاده کنید:

a. (۳ نمره) مقدار C حاصل را برای مجموعه داده‌ی مربوطه مشخص نمایید. این مقدار منجر به چه درصدی از خطای روی نمونه‌های اعتبارسنجی می‌شود.

b. (۲ نمره) درصد خطای روی نمونه‌های آزمون را در حالتی که C انتخاب شده در قسمت a در SVM استفاده شده باشد، مشخص نمایید. همچنین نتایج به دست آمده برای درصد خطای روی نمونه‌های اعتبارسنجی و آزمون را مورد مقایسه و بحث قرار دهید.