دانشگاه تهران
دانشکدگان فارابی
دانشکده مهندسی

تکلیف شماره‌ی ۵

تکلیف پنجم

انتخاب ویژگی و تولید ویژگی

FEATURE SELECTION AND FEATURE GENERATION

۱) اگر x_i و y_i، $i=1,2,\ldots,N$ نمونه‌هایی مستقل از دو توزیع گاوسی با واریانس مشابه σ^2 باشد، نشان دهید که متغیر تصادفی

$$\frac{(2N-2)s_z^2}{\sigma^2}$$

که در آن

$$s_z^2=\frac{1}{2N-2}\left(\sum_{i=1}^{N}(x_i-\bar{x})^2+\sum_{i=1}^{N}(y_i-\bar{y})^2\right)$$

که در آن $\bar{x}$ و $\bar{y}$ مقادیر میانگین نمونه‌ای متناظر هستند، دارای توزیع مربع خی (chi-square) با $2N-2$ درجه‌ی آزادی است.

۲) فرض کنید N_1 و N_2 به ترتیب مقادیر موجود یک ویژگی در طبقه ۱ و طبقه‌ی ۲ باشد. فرض می‌شود که این ویژگی از یک توزیع گاوسی با واریانس مشابه در هر دو طبقه پیروی کند. آماره‌ی آزمون را به صورت

$$q=\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{s_z\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}}$$

تعریف می‌کنیم که در آن

$$s_z^2=\frac{1}{N_1+N_2-2}\left(\sum_{i=1}^{N_1}(x_i-\bar{x})^2+\sum_{i=1}^{N_2}(y_i-\bar{y})^2\right)$$

و m_1 و m_2 به ترتیب، مقادیر میانگین واقعی است. نشان دهید که q از توزیع t با درجه‌ی N_1+N_2-2 آزادی پیروی می‌کند.

۳) نشان دهید اگر متغیرهای $x_i, i=1,2,\ldots,l$ دارای توزیع توأم گاوسی باشند، آن‌گاه متغیرهای $y_i, i=1,2,\ldots,l$ حاصل یک تبدیل خطی از x_iها نیز دارای توزیع توأم گاوسی خواهد بود. به علاوه اگر x_iها دو به دو مستقل باشند و تبدیل متعامد باشد، آن‌گاه y_iها نیز دو به دو مستقل و گاوسی خواهد بود.

۴) نشان دهید که اگر ویژگی‌ها مستقل آماری باشند، آن‌گاه واگرایی طبقه‌ها (divergence) با رابطه‌ی زیر داده می‌شود:

$$d_{ij}(x_1,x_2,\ldots,x_l)=\sum_{k=1}^{l}d_{ij}(x_k).$$

۵) نشان دهید ماتریس پراکندگی مخلوط (mixture scatter) مجموع ماتریس‌های پراکندگی درون‌طبقه‌بندی (within-class scatter) و میان‌طبقه‌ای (between-class scatter) است.

۶) نشان دهید که برای یک مسئله‌ی دوطبقه‌ای، مقدار ویژه‌ی غیر صفر ماتریس $S_w^{-1}S_b$ برابر است با

$$\lambda_1=P_1P_2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^TS_{xw}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$$

و بردار ویژه‌ی متناظر

$$\mathbf{v}_1=S_{xw}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$$

است که در آن P_1 و P_2 احتمالات طبقه‌ای متناظر است.

۷) نشان دهید که اگر ماتریس‌های $S_۱$ و $S_۲$ دو ماتریس کوواریانس باشد، آن‌گاه بردارهای ویژه‌ی $S_۱^{-۱}S_۲$ نسبت به $S_۱$ عمود هستند، یعنی

$$\mathbf{v}_i^T S_۱ \mathbf{v}_j = \delta_{ij}$$

راهنمایی: از این واقعیت استفاده کنید که $S_۱$ و $S_۲$ می‌توانند به صورت همزمان قطری‌سازی شوند (پیوست B کتاب درسی را ببینید).

۸) برای ماتریس زیر، بازنمایی SVD را محاسبه کنید.

$$\begin{bmatrix} ۱ & ۲ \\ ۲ & ۱ \\ ۱ & ۳ \end{bmatrix}$$

۹) یک برنامه‌ی MATLAB بنویسید که مؤلفه‌های اصلی ماتریس کوواریانس یک ماتریس X از داده‌ها با ابعاد $l \times N$ را به همراه واریانس‌های متناظر آنها محاسبه کند (PCA).

مسئله‌هایی که در کنار آنها نماد درج شده است، برای حل نیاز به برنامه‌نویسی کامپیوتری (ترجیحاً در محیط MATLAB) دارند. برای تحویل، برنامه‌ها به همراه گزارش نتایج در محل مشخص شده در سایت در قالب یک فایل آرشیو `zip.` آپلود شود.