Assignment 1

1 Generative and Discriminative classifiers: Gaussian Bayes and Logistic Regression

Recall that a generative classifier estimates $P(\mathbf{x}, y) = P(y)P(\mathbf{x}|y)$, while a discriminative classifier directly estimates $P(y|\mathbf{x})$.

1.1 Specific Gaussian naive Bayes classifiers and logistic regression

Consider **a specific class** of Gaussian naive Bayes classifiers where:

- y is a boolean variable following a Bernoulli distribution, with parameter $\pi = P(y = 1)$ and thus $P(Y = 0) = 1 - \pi$.
- $\mathbf{x} = [x_1, \ldots, x_D]^T$, with each feature x_i a continuous random variable. For each x_i, $P(x_i|y = k)$ is a Gaussian distribution $\mathcal{N}(\mu_{ik}, \sigma_i)$. Note that σ_i is the standard deviation of the Gaussian distribution, which does not depend on k.
- For all $i \neq j$, x_i and x_j are conditionally independent given y (so called "naive" classifier).

Question: please show that the relationship between a discriminative classifier (say logistic regression) and the above specific class of Gaussian naive Bayes classifiers is precisely the form used by logistic regression.

1.2 General Gaussian naive Bayes classifiers and logistic regression

Removing the assumption that the standard deviation σ_i of $P(x_i|y=k)$ does not depend on k. That is , for each x_i, $P(x_i|y=k)$ is a Gaussian distribution $\mathcal{N}(\mu_{ik}, \sigma_{ik})$, where $i = 1, \ldots, D$ and $k = 0, 1$.

Question: is the new form of $P(y|\mathbf{x})$ implied by this more general Gaussian naive Bayes classifier still the form used by logistic regression? Derive the new form of $P(y|\mathbf{x})$ to prove your answer.

1.3 Gaussian Bayes classifiers and logistic regression

Now, consider the following assumptions for our Gaussian Bayes classifiers (without "naive"):

- y is a boolean variable following a Bernoulli distribution, with parameter $\pi = P(y=1)$ and thus $P(Y=0) = 1-\pi$.
- $\mathbf{x} = [x_1, x_2]^T$, i.e., we only consider two features for each sample, with each feature a continuous random variable. x_1 and x_2 are **not** conditional independent given y. We assume $P(x_1, x_2|y=k)$ is a bivariate Gaussian distribution $\mathcal{N}(\mu_{1k}, \mu_{2k}, \sigma_1, \sigma_2, \rho)$, where μ_{1k} and μ_{2k} are means of x_1 and x_2, σ_1 and σ_2 are standard deviations of x_1 and x_2, and ρ is the correlation between x_1 and x_2. The density of the bivariate Gaussian distribution is:

$$P(x_1, x_2|y=k) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp[-\frac{\sigma_2^2(x_1-\mu_{1k})^2 + \sigma_1^2(x_2-\mu_{2k})^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_{1k})(x_2-\mu_{2k})}{2(1-\rho^2)\sigma_1^2\sigma_2^2}].$$

Question: is the form of $P(y|\mathbf{x})$ implied by such not-so-naive Gaussian Bayes classifiers still the form used by logistic regression? Derive the form of $P(y|\mathbf{x})$ to prove your answer.