E1 213 Pattern Recognition and Neural networks

Problem Sheet 4 – Answer to Q1

1. Consider a general K-class problem with a general loss function. Let $h(X)$ denote the output of the classifier on X. Let $R(\alpha_i|X)$ denote the expected loss when classifier says α_i and conditioned on X. That is, $R(\alpha_i|X) = E[L(h(X), y(X))|h(X) = \alpha_i, X]$, where, as usual, $y(X)$ denotes the 'true class'. We had only considered deterministic classifiers where h is a function that assigns a unique class label for any given X. Suppose we use a stochastic classifier, h, which, given X, outputs α_i with probability $p_h(\alpha_i|X)$. (Note that we would have $p_h(\alpha_i|X) \geq 0$ and $\sum_i p_h(\alpha_i|X) = 1$). For this classifier, show that the risk is given by

$$R(h) = \int \left[\sum_{i=1}^{K} R(\alpha_i|X) p_h(\alpha_i|X) \right] f(X)\, dX$$

where $f(X)$ is the density of X. Using the above expression, find the best choice of values for all the $p_h(\alpha_i|X)$ and hence conclude that we do not gain anything by making the classifier stochastic.

Answer

Using the same notation as in class, we have

$$\begin{aligned}
R(h) &= E[E[L(h(X), y(X)) \mid X]] \\
&= E[\sum_{i=1}^{K} (E[L(h(X), y(X)) \mid h(X) = \alpha_i, X] \Pr[h(X) = \alpha_i \mid X]] \\
&= E[\sum_{i=1}^{K} R(\alpha_i|X) p_h(\alpha_i|X)] \\
&= \int \left[\sum_{i=1}^{K} R(\alpha_i \mid X) p_h(\alpha_i|X) \right] f(X)\, dX
\end{aligned}$$

Suppose for some X, we have $R(\alpha_i \mid X) \leq R(\alpha_j \mid X)$, $\forall j$. Since p_h is a probability mass function, we have for any classifier h

$$\min_j R(\alpha_j \mid X) \leq \sum_{i=1}^{K} R(\alpha_i \mid X) p_h(\alpha_i|X)$$

Let $h2$ be any stochastic classifier with $0 < p_{h2}(\alpha_i \mid X) < 1$. Suppose $h1$ be a dterministic classifier with $h1(X) = \alpha_i$. Then we have

$$R(h1(X) \mid X) = R(\alpha_i \mid X) < \sum_{i=1}^{K} R(\alpha_i \mid X) p_{h2}(\alpha_i|X)$$

Thus, a deterministic classifier that, at every X, chooses class label to minimize risk at that X has lower risk than any stochastic classifier, proving that Bayes classifier would be a deterministic classifier.