Bayesian Decision theory with continuous features

- Consider more than one feature through feature victor **x,** where **x** is in d-dimensional Euclidean space R^d, called the feature space.
- Consider c (more than two) states of nature as in $\{w_1, \ldots, w_c\}$.
- Consider a actions, more than merely deciding the state of nature $\{\alpha_1, \ldots, \alpha_a\}$
- Introduce a loss function, more general than the probability of error, $\lambda(\alpha_i|w_j)$, describing the loss incurred for taking action α_i when the state of nature is w_j.

$P(w_j|\mathbf{x}) = p(\mathbf{x}|w_j) * P(w_{j'}) / p(\mathbf{x})$ where he evidence is now

where the evidence is now

$p(\mathbf{x}) = \Sigma\, p(\mathbf{x}|w_j) * P(w_j)$, $j = 1,\ldots,c$.

If we have observed a particular **x** and we intend to take action α_i but the true state of nature is w_j then we incur loss $\lambda(\alpha_i|w_j)$.

Because $P(w_j|\mathbf{x})$ is the probability that the true state of nature is w_j, the expected loss associated with taking action $\alpha_{i\text{ is}}$

$R(\alpha_i|\mathbf{x}) = \Sigma\, \lambda(\alpha_i|w_j) * P(w_j|\mathbf{x})$, $j = 1,\ldots,c$.

It is called the risk and our objective is to minimize the risk. A general decision rule is a function $\alpha(\mathbf{x})$. For every **x,** the decision function $\alpha(\mathbf{x})$ assumes one of the a values $\alpha_1,\ldots,\alpha_a$. The overall risk is then given by

$R = \int R(\alpha(\mathbf{x})|\mathbf{x})p(\mathbf{x})\, d\mathbf{x}$.

To minimize the overall risk, we compute the risk

$R(\alpha_i|\mathbf{x}) = \Sigma\, \lambda(\alpha_i|w_j) * P(w_j|\mathbf{x})$, $j = 1,\ldots,c$ and select the action α_i for which $R(\alpha_i|\mathbf{x})$ is minimum. The resulting minimum overall risk is called the Bayes risk and is the best performance that can be achieved.

Special case of two-category classification problem:

Here, action α_1 corresponds to deciding that the true state of nature is w_1 and action α_2 corresponds to deciding that it is w_2.

Simplifying, write λ_{ij} for $\lambda(\alpha_i|w_j)$, the loss incurred for deciding w_i when the true nature is w_j.

So, $R(\alpha_1|x) = \lambda_{11}P(w_1|x) + \lambda_{12}P(w_2|x)$ and

$R(\alpha_2|x) = \lambda_{21}P(w_1|x) + \lambda_{22}P(w_2|x)$.

Our rule would be to decide w_1 if $R(\alpha_1|x) < R(\alpha_2|x)$.

That is, in terms of posterior probabilities, we decide w_1 if $(\lambda_{21} - \lambda_{11})\, P(w_1|x) > (\lambda_{12} - \lambda_{22})\, P(w_2|x)$.

This can also be expressed in terms of prior probabilities, by Bayes formula, as, decide w_1 if

$(\lambda_{21} - \lambda_{11})p(x|w_1)P(w_1) > (\lambda_{12} - \lambda_{22})p(x|w_2)P(w_2)$ else decide w_2.