Lecture 5 – unannotated slides

Minimax criteria

- Sometimes we must design our classifier to perform well over a range of prior probabilities.
- For instance, the likelihood may remain same, but the prior may completely change in a new setup. Or we want to use a classifier in a new setup and prior is unavailable.

- Approach – minimize maximum possible overall risk

- In order to understand this, we let R1 denote that (as yet unknown) region in feature space where the classifier decides ω1 and likewise for R2 and ω2, and then write our overall risk

$$R = \int_{R_1} [\lambda 11 P\,(\omega 1)\; p(x|\omega 1) + \lambda 12 P\,(\omega 2)\; p(x|\omega 2)]\; dx \;+$$

$$\int_{R2} [\lambda 21 P\,(\omega 1)\; p(x|\omega 1) + \lambda 22 P\,(\omega 2)\; p(x|\omega 2)]\; dx\,.$$

- We use the fact that P(ω2) = 1 − P(ω1) and that

$$\int_{R1} p(x|\omega 1)\, dx + \int_{R2} p(x|\omega 1)\, dx = 1$$

$$R(P(\omega 1)) = \underbrace{\lambda 22 + (\lambda 12 - \lambda 22)\, p(x|\omega 2) dx}_{\text{Minimax risk}} +$$

$$P\,(\omega 1)\underbrace{\left[(\lambda 11 - \lambda 22) - (\lambda 21 - \lambda 11) \int_{R2} p(x|\omega 1)\, dx - (\lambda 12 - \lambda 22) \int_{R1} p(x|\omega 2)\, dx\,\right]}_{\text{0 for minimax sol}}$$

Classifiers, Discriminant Functions and Decision Surfaces

- The multi-category case (consider zero-one loss only)
  - Set of discriminant functions $g_i(x), i = 1,\ldots, c$
  - The classifier assigns a feature vector x to class ω_i if:

$$g_i(x) > g_j(x) \quad \forall j \neq i$$

- $Let\ g_i(x) = -R(\alpha_i \mid x) = P(\omega_i \mid x) - 1$

  (max. discriminant corresponds to min. risk!)

- For the minimum error rate, we take

$$g_i(x) = P(\omega_i \mid x)$$

  (max. discrimination corresponds to max. posterior!)

$$g_i(x) \equiv P(x \mid \omega_i)\, P(\omega_i)$$

$$g_i(x) = \ln P(x \mid \omega_i) + \ln P(\omega_i)$$

(ln: natural logarithm!)