1 Problem1

1.1 1a

Let's consider two elements a and b, then the decision boundary for element x would be equivalent to:

$$||a-x||^2 = ||b-x||^2$$
$$a^2 - 2(a,x) = b^2 - 2(b,x)$$

which corresponds to linear decision boundary.

1.2 1b

Let's consider a class A, then for the element x the decision rule of belonging x to A would be:

$$\exists a \in A \quad \forall b \notin A$$
$$||a-x||^2 \leq ||b-x||^2$$
$$a^2 - 2(a,x) \leq b^2 - 2(b,x)$$

It would be the same for other $N-1$ classes. So, we will receive intersection of piecewise linear curves.

2 Problem2

2.1 2a

$$cost(\hat{w_k}, w_i) = \begin{cases} 0 & k = i \\ \lambda_i & else \end{cases}$$

- Expected loss of prediction $\hat{w_i}$

$$L(\hat{w_i}) = \sum_{j \neq i}^{C} \lambda_j p(w_j|x) = \sum_{j \neq i}^{C} \lambda_j \frac{p(w_j)p(x|w_j)}{p(x)}$$

- Bayes dicision rule minimizes expected loss

$$\hat{w*} = \underset{\hat{w}}{\text{argmin}} L(\hat{w})$$

2.2 2b

$$L(w_i) = \sum_{j \neq i} \lambda p(w_j|x) = \lambda \sum_{j \neq i} p(w_j|x) = \lambda(1 - p(w_i|x))$$
$$\hat{w*} = \underset{\hat{w}}{\text{argmin}} L(\hat{w})$$
$$\hat{w*} = \underset{\hat{w}}{\text{argmax}} p(\hat{w}|x)$$

3 Problem3

3.1 3a

Let's consider node t, N_t - number of elements at node t.
For one feature and one node t: we should do * N_t operations with fixed threshold element to calculate probabilities of classes within the node t;
* do it N_t times.
For D features we have DN_t^2. Then let's calculate it for all nodes in the tree. If we unify nodes by levels, at each level

$$\sum_{t \in level} N_t = N$$

. Then

$$D \sum_{t \in level} N_t^2 \leq D \sum_{t \in level} NN_t = DN^2$$

. The number of levels is equal to $log_2 N$, so the result $O(DN^2 logN)$

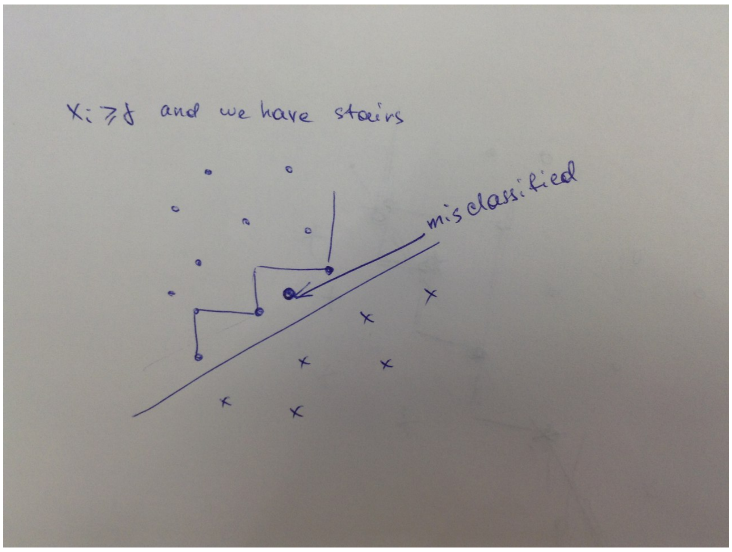
3.2 3b

Let's sort features values among sample N_t. The cost is equal to $DNlogN$. Then the probabilities of classes within the node t will be calculated in $O(1)$, because we have a fully sorted array. Then, calculating the sum for all levels will be the same as in previous subproblem, so we have $O(DNlog^2 N)$

4 Problem4

4.1 4a

The example of misclassification can be found in a picture. Using binary tree we will have stairs as a result of making decision at each node, when a linear separability is right.



. 1: example of misclassification