homework sheet 02

Marco Seravalli
03626387
`marco.seravalli@tum.de`

Nikola Tchipev
03625168
`n.tchipev@tum.de`

1 Decision Trees

Consider the Voronoi diagram generated by the given points. Color every Voronoi region which was generated around a point labeled a in white and otherwise in black. Then the problem of finding a decision tree, which correctly labels all n points is (clearly) equivalent to the problem of determining the color at a random position inside the unit square.

We will work the second formulation.

Problem 1

Claim: the worst case scenario, demanding most splits in the decision tree, is when we have a "checkerboard" coloring - when no two Voronoi regions of the same colour share an edge (see Fig. 1). Any other scenario is easier or not-more-difficult to solve.

(informal) Proof: consider a Voronoi tesselation in which there exist two regions of the same color, which share an edge. Then if we merge those two regions, we can remove one of the points. This reduces the number of regions to search for by one, which cannot make the problem more complicated, i.e. no scenario is more complicated than the "checkerboard" coloring.

With this statement, it is enough to consider the "checkerboard" coloring and it's solution tree:

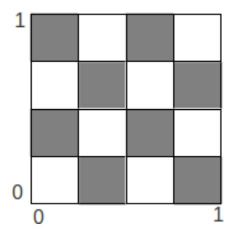

Figure 1: Checkerboard colouring for n=16

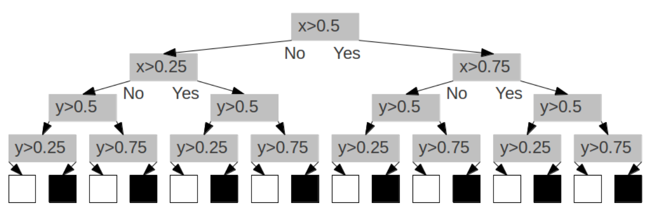

Figure 2: Decision tree for the scheme from Fig.1

Alternative argument - "quadtree"-like approach:
Suppose that the domain contains n points. Then we can always split it in two s.t. each subdomain contains approximately the same number of points (to be precise, each subdomain contains at most $\frac{2^{\lceil \log_2 n \rceil}}{2}$ points). Apply this recursively to each subdomain. We get a tree of depth at most $\lceil \log_2 n \rceil$.

Problem 2

Our worst case scenario in the solution of Problem 1 fulfills the requirements as the tree is of depth $n-1$.

Problem 3

For this problem, we switch back to the original formulation with the classification of the points. We will still consider them as black and white, however.

We present the solution in Fig. 2 and 3.

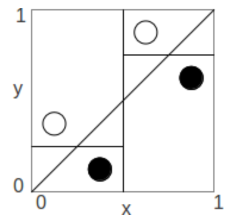

Figure 3: Configuration, separable by a line, demanding $n-1$ splits

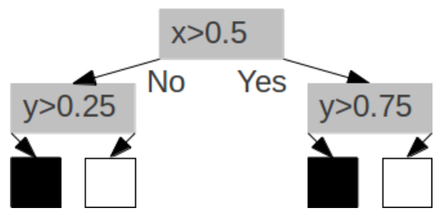

Figure 4: Decision tree, corresponding to point configuration from Fig.3

2 Nearest Neighbours

Problem 4

Rewrite the distance

$$d(\mathbf{x},\mathbf{y}) = \Sigma_i \sigma_i (\mathbf{x}_i - \mathbf{y}_j)^2,\ \sigma_i > 0$$

in the form of *Mahalanobis distance*. Hence $d(\mathbf{x},\mathbf{y})$ becomes:

$$d(\mathbf{x},\mathbf{y}) = (\mathbf{x}-\mathbf{y})^T \Sigma (\mathbf{x}-\mathbf{y})$$

where Σ is defined as a symmetric square matrix

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & & \end{bmatrix}$$

Problem 5

In the Euclidean distance function if the scale of one feature dominates the others this will be accentuated, so that the other features will become insignificant. A possible solution could be to precondition the problem by scaling the feature that outweighs the other ones. For example if the scale of the j^{th} feature is 1000 times larger than the others a possible solution could be to modify the corresponding σ_j of the matrix Σ

$$\sigma_j = \frac{1}{1000}$$

so that the new feature has a weight that is closer to the other dimensions.

Another possibility could be to set Σ as the covariance matrix.

Problem 6

Show that

$$\frac{p(c=0|x^*)}{p(c=1|x^*)} \approx \frac{exp((-\|x^*-x_0\|)^2/(2\sigma^2))}{exp((-\|x^*-x_1\|)^2/(2\sigma^2))}$$

for $\sigma \to 0$ corresponds to the kNN for $k=1$.

We have the probabilities

$$p(c=0) = \frac{N_0}{N_0+N_1}$$

$$p(c=1) = \frac{N_1}{N_0+N_1}$$

Then

$$\frac{p(c=0|x^*)}{p(c=1|x^*)} = \frac{p(x^*|c=0)p(c=0)}{p(x^*|c=1)p(c=1)} \Rightarrow$$

$$\frac{p(c=0|x^*)}{p(c=1|x^*)} = \frac{\frac{N_0}{N_0+N_1}\frac{1}{N_0(2\Pi\sigma^2)}\Sigma_{n\in class0} exp((-\|x^*-x_0\|)^2/(2\sigma^2))}{\frac{N_1}{N_0+N_1}\frac{1}{N_1(2\Pi\sigma^2)}\Sigma_{n\in class1} exp((-\|x^*-x_1\|)^2/(2\sigma^2))} \Rightarrow$$

$$\frac{p(c=0|x^*)}{p(c=1|x^*)} = \frac{\Sigma_{n\in class0} exp((-\|x^*-x_0\|)^2/(2\sigma^2))}{\Sigma_{n\in class1} exp((-\|x^*-x_1\|)^2/(2\sigma^2))}$$

This last equation is dominated by the nearest point to x^* so we obtain

$$\frac{p(c=0|x^*)}{p(c=1|x^*)} \approx \frac{exp((-\|x^*-x_0\|)^2/(2\sigma^2))}{exp((-\|x^*-x_1\|)^2/(2\sigma^2))} \Rightarrow$$

$$\frac{p(c=0|x^*)}{p(c=1|x^*)} \approx exp\frac{(-\|x^*-x_0\|)^2-(-\|x^*-x_1\|)^2}{(2\sigma^2))}$$

In such case if $(\|x^*-x_0\|) < (\|x^*-x_1\|)$ we have that

$$\frac{p(c=0|x^*)}{p(c=1|x^*)} = \infty$$

which means that the x^* would be marked as belonging to class 0.

If instead we have $(\|x^*-x_0\|) > (\|x^*-x_1\|)$ this yields to

$$\frac{p(c=0|x^*)}{p(c=1|x^*)} = 0$$

that corresponds as marking the new x^* to class 1. This is the very same principle as the kNN, i.e. assigning the new point to the same class of the closest neighbour.

3 Random Projections

Problem 7

Consider the 1D situation. Then there are only two possible cases:

- the vectors $\mathbf{u}$ and $\mathbf{v}$ are parallel
  Then $h_r(\mathbf{u}) = h_r(\mathbf{v})$ is always fulfiled and
  $$p(h_r(\mathbf{u}) = h_r(\mathbf{v})) = 1 - \frac{0}{\pi} = 1$$
- the vectors $\mathbf{u}$ and $\mathbf{v}$ are antiparallel
  Then $h_r(\mathbf{u}) = -h_r(\mathbf{v})$ and
  $$p(h_r(\mathbf{u}) = h_r(\mathbf{v})) = 1 - \frac{\pi}{\pi} = 0$$

Now suppose we are in 2D. It is sufficient to show that

$$p\left(h_r(\mathbf{u}) \neq h_r(\mathbf{v})\right) = \frac{\theta(\mathbf{u},\mathbf{v})}{\pi}.$$

Then

$$p\left(h_r(\mathbf{u}) = h_r(\mathbf{v})\right) = 1 - p\left(h_r(\mathbf{u}) \neq h_r(\mathbf{v})\right) = 1 - \frac{\theta(\mathbf{u},\mathbf{v})}{\pi}.$$

Let $u_\perp$, $v_\perp$ be the lines, perpendicular to $\mathbf{u}$ and $\mathbf{v}$, respectively. Then the probability that $h_r(\mathbf{u})$ is not equal to $h_r(\mathbf{u})$ is the same as the probability of choosing any point in the plane, which falls in the shaded region in Fig. 5. From simple geometry, the angle between $\mathbf{u}_\perp$ and $\mathbf{v}_\perp$ is equal to $\theta(\mathbf{u},\mathbf{v})$. Thus the probability that a randomly chosen point in the plane falls in the shaded region is equal to $\frac{2\theta}{2\pi} = \frac{\theta}{\pi}$, which we wanted to show.

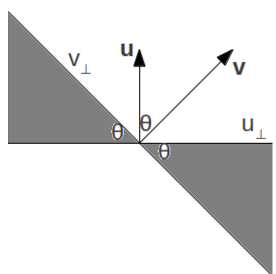

Figure 5: 2D diagram

Now suppose that we are in a higher dimensional space. Then if $\mathbf{u}$ and $\mathbf{v}$ are not parallel or antiparallel (in which case we come down to the same calculation as in 1D, which we already showed), they define a unique plane that contains both of them. It is sufficient to consider the projection of $\mathbf{r}$ onto that plane and whether it lies in the same shaded region as in 2D, hence the problem is reduced to the 2D case.