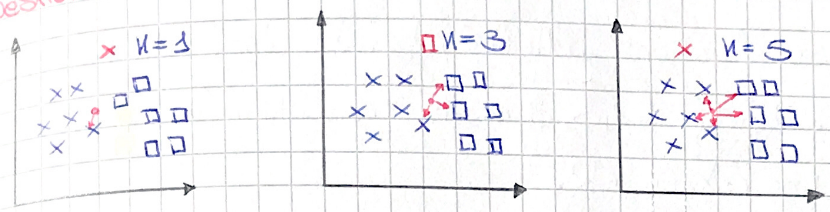
EXAM TEST 2 2017/18

Question 1



In the first and the third case the new instance will be classified as x, in the second case as □.

Question 2

- Convolutional stage: in this stage is performed the convolution operation

$I * K(i,j) = \sum_{m \in S_1} \sum_{n \in S_2} I(m,n)\, K(i-m, j-n)$.

In this stage are applied two principles:

* Sparse connettivity: each node is not connected to all the nodes of the successive layer but only to the closest.
* Parameter sharing: we force some parameters to have the same value. In this mode we reduce the number of trained parameters.

- Detector stage: non-linear activation function (Relu, Sigmoid).
- Pooling stage: we can have max-pooling or average-pooling. In max pooling we compute the max of a rectangular region, in average pooling the average.

Question 3

We want to maximize the data variance after the projection to some direction u_1.

Projected points: $u_1^T x_n$. We want to maximize the projected variance:

$\max_{u_1} u_1^T S u_1$ s.t. $u_1^T u_1 = 1$ This is equivalent to

$\max_{u_1} u_1^T S u_1 + \lambda_1 (1 - u_1^T u_1)$. If we set the derivative to 0:

$S u_1 = \lambda_1 u_1$ u_1 must be an eigenvector of S.

$u_1^T S u_1 = \lambda_1$ is the variance after the projection.

The variance is maximum if u_1 is the eigenvector corresponding to the largest eigenvalue λ_1. This is called principal component.

Question 4

We can use a MDP $= < X, A, \delta, r >$ with X set of states and $X = \{ls, rs\}$ left and right side of the river, A set of actions with $A = \{br_1, br_2, br_3\}$ that are the 3 possible actions: traverse bridge 1, 2, 3. δ is the transition function $\delta: X \times A \to X$ and r is the reward function $r: X \times A \to \mathbb{R}$.

For this problem we can use Q-Learning:

For each x we initilize $\hat{Q}_{(0)}(x, a) \leftarrow 0$

We observe the current state;

for each time $t = 1 \dots T$

We choose an action and we execute this action;

We collect the reward and we observe the new state;

$$\hat{Q}_{(t)}(x, a) \leftarrow \bar{r} + \gamma \max_{a' \in A} \hat{Q}_{(t-1)}(x', a')$$

The optimal policy is $\pi^*(x) = \underset{a \in A}{\operatorname{argmax}} \hat{Q}_{(T)}(x, a)$

We can use the ε-greedy policy for balancing exploration and exploitation: we choose exploration with probability ε and exploitation with probability $1 - \varepsilon$.

Question 5

1). Given a classification problem with K classes is possible to compute, using a confusion matrix, how many times an instance of a class C_i is classified into a class C_j.

2).

	LOW	MED	HIGH
LOW	20	20	10
MED	5	10	15
HIGH	5	5	40

3). The accuracy is the sum of the elements in the main diagonal divided by the sum of all the elements.

$$\text{accuracy} = \frac{20 + 10 + 40}{150} = \frac{7}{15} \simeq 50\%$$

Question 6

1). $E(w) = \frac{1}{2} \sum_{n=1}^{N} (t_n - w^T \phi(x_n))$ is the error function.

2). $\phi_4(x, y) = xy \quad \phi_5(x) = x^2 \Longrightarrow y(X) = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 \phi_4(x_1, x_2) + w_5 \phi_5(x_3)$

3). We can use sequential learning: $\hat{w} \leftarrow \hat{w} + \eta [t_n - w^T \phi(x_n)] \phi(x_n)$