**Folie1:**

**Folie2:**

**Folie3:**

**This is my workflow.**

**First of all, I have cleaned the training data. Meaning that I have removed all columns that contain only zeros or ones.**

**Second: I have randomally created an initial population, that is a binary population.**

**In the third step, I have used the F1 score to evaluate each individual of the intial population and each row in the training data.**

**In the fourth step, I have used selection to select only some of the individuals from the previous step.**

**Afterthat, I used the uniform crossover and the swap mutation in order to insure obtaining a general solution.**

**Steps 3 until 6 will be repeated until all iteration are done.**

**After the last iteration, the best parameters are selected.**

**These will be trained 10 times. By that it is meant that step 3 until 6 will be applied for each parameter. This can be seen in yellow**

**From the results of this step, I choose the best classifierbased on F1 score.**

**In the last step, I test this classifier on the test data.**

**Folie 4:**

**The initial population is generated randomally. The size of the population ranges between 50-300 individuals. miRNAs are chosen randomaly and their values can be either 0 or 1. These are three examples of intitial population.**

**Folie 5:**

**I have chosen the F1 score to evaluate my function. The formula of the F1 score can be seen in this slide. Each individual will be evaluated with each row of the training data.**

**In order to compute the F1 score we have to consider two cases.**

**First case: annotation is 0. We should first compute the match and unmatch values in each position [I will explain this in the next slide]. If unmatch bigger than or equal to match then TN will be 1, otherwise FP will be 1.**

**Second case: annotation is 1. We compute the match and unmatch similar to the previous case. If unmatch less than or equal to match then TP will be 1, otherwise FN will be 1.**

**At the end, individuals are sorted in descending order based on their score.**

**Folie 6:**

**I will show you a simple example on how the F1 score is computed.**

**--CLICK--**

**Imagine we have a random population with only 2 individuals**

**--CLICK--**

**And imagine we have a trainng data with 3 rows. and the annotations for these rows are 0, 1, 0.**

**--CLICK--**

**In order to compute the F1 score, we have to compare individual 1 and row 1. By that I mean we have to calculate the match and unmatch values for each position.**

**--CLICK--**

**Because annots is 0, match is 3 and unmatch is 2, FP will be 1, as explain in the previous slide.**

**--CLICK--**

**Now we compare individual 1 and row 2.**

**--CLICK--**

**In this case we have annots is 1, match is 3 and unmatch is 2. Thus, TP is 1.**

**--CLICK--**

**We do the same but now we compare individual 1 and row 3.**

**--CLICK--**

**Here, FP will be 1 because annots is 0, and unmatch is begger than match**

**--CLICK--**

**As we can see, we have two values of FP, so how can we compute the F1 score?**

**--CLICK--**

**--CLICK--**

**For that we need to sum up all TPs, FPs and FNs values, and then we can compute the F1 sore between individual 1 and all rows in the training data using the F1 formula shown in the previous slide.**

**--CLICK--**

**--CLICK--**

**--CLICK--**

**We do the same when we want to compute the F1 score for all other individuals.**

**Folie 7:**

**As mentioned before, the crossover population will be sorted in descending order, based on their F1 score. For example: individual 1 has the highest score, while individual 8 has the lowest score.**

**--CLICK--**

**Becaue it is sorted, I selected the first half of this population.**

**--CLICK--**

**This is my new crossover population. we each individual is called parent.**

**Folie 8:**

**Uniform crossover is applied to each subsequent parent. Uniform crossover generates a random positions of the parent pairs. The elements at the**

**randomly generated positions of both parents are swapped with each other. This creates two new individuals called children.**

**--CLICK--**

**As an example: let 0, 1 and 5 be the randomely generated positions.**

**--CLICK--**

**We consider these two parents**

**--CLICK--**

**--CLICK--**

**--CLICK--**

**Here we swap the elements (miRNAs) in positions 0, 1 and 5**

**--CLICK--**

**--CLICK--**

**--CLICK--**

**As a results, we get two new children, where the values in positions 0, 1, and 5 are swapped.**

**--CLICK--**

**After computing all uniform crossover between all parents, we will get a new**

**population that has the same size as the initial population. We pack first all parents togerther then we pack all children together.**

**--CLICK--**

**--CLICK--**

**I call this as mutation population.**

**Folie 9:**

**Folie 10:**

**I have chosen the same parameter ranges as we have in the lecture.**

**I have randomally generated 100 parameters combinations, and then I chose the best parameters combination based on the F1 score. These I have trained 10 times to find my perfect classifier. This classifier I used for testing my test data.**

**-----------------------------------**

**IN FOLIEN SCHREIBEN:**

**Conculstion:**

**Liver\_data\_1:**

**- the data should contain at least 6 miRNAs**

**- Mutation probability, Number of Iteration and Tausch have no great effect on the F1 score**

**- It is good if the population size is bigger than 100**

**- hsa-miR-106a and hsa-miR-221 are always present.**

**-Hsa-miR-106 is almost always up- regulated, and hsa- miR-221 is almost always down-regulated.**

**Liver\_data\_2:**

**- the data should contain at least 7 miRNAs**

**- Mutation probability, Number of Iteration, Population size and Tausch have no great effect on the F1 score.**

**- Since there are so many different miRNAs that are randomly selected, it is difficult to decide which miRNA predicts very well.**

**- But with the 10 tests you can conclude that a down-regulated**

**hsa-miR-199a\* will give a good prediction.**