Introduction to Machine Learning for Physicists

Problem Set 4
Lecturer: Fabian Ruehle
Github: https://github.com/ruehlef/ML_Oxford_Hilary2020/Problem_set_4/

4.1 Comparing clustering algorithms

In this exercise, we will compare different clustering algorithms on the same input set. Concretely, we want to compare

- K-Means
- Mean shift
- DBSCAN
- Birch

We will use again the data from the $Z \to e\bar{e}$ decay of Exercise 1.2. The library sklearn implements all these algorithms (and many more). The documentation (as well as a nice comparison and an explanation of the various algorithms) can be found at https://scikit-learn.org/stable/modules/clustering.html. I provide again a template and solution file.

(a) Plot the data points in the ($\Delta\eta$, $\Delta\Phi$)-plane. Remember that $\eta_{1,2}$ are the pseudo-rapidities and $\phi_{1,2}$ are the azimuthal angles. What do you observe? What is the physical reason for the distribution. Would it look different at a different collider?

(b) Perform K means clustering. Where are the cluster centers? What does that mean physically?

(c) Perform Mean shift clustering. Where are the cluster centers? What does that mean physically?

(d) Perform DBSCAN clustering. Check how different parameters change the result.

(e) Perform Birch clustering.

4.2 Genetic algorithms

In this exercise we will minimize a scalar potential of an effective 4D $\mathcal{N} = 1$ theory as it might arise e.g. in a toy example of string theory (we will not go into a swampland discussion here, but just use the typical form of the potential and some semi-realistic parameters; the point of the exercise is to illustrate GAs). While there are libraries for Genetic Algorithms available, we will implement it from scratch since it is very easy and instructive to do so. I provide again a template and solution file.
Typically, there will be a dilaton S (which controls the coupling strength) and a volume modulus T (which controls the size of the extra dimensions). The theory will have a superpotential

$$W = W_0 + Ae^{-aS} + Be^{-bT}\,, \tag{1}$$

and a Kähler potential

$$K = -\log(S + S^*) - 3\log(T + T^*)\,. \tag{2}$$

Since S and T are moduli, they only have non-perturbative superpotential terms. Furthermore, the axions (the imaginary parts of S and T) drop out of the Kähler potential, and their shift symmetry is broken by the non-perturbative effects. The coefficients W_0, A, a, B, b depend on the theory (on vacuum expectation values of other fields, Pfaffians and determinants, beta functions, ...). We will, for the sake of this example, set $W_0 = -1$, $A = 2$ $B = 10$, $a = 1$ and $b = \pi$. We denote the scalar components of S and T by $s + i\eta$ and $t + i\tau$, respectively. We can then write the scalar potential V as

$$\begin{aligned} V = -\frac{1}{16st^3}\big(4 + e^{-2s}\left(4s^2 + 4s + 1\right) - 4e^{-s}(2s+1)\cos(\eta) \\ + 3e^{-2\pi t}\left(4\pi^2 t^2 + 12\pi t + 3\right) - 12e^{-\pi t}(2\pi t + 1)\cos(\tau) \\ + 6e^{-s-\pi t}(2s + 2\pi t + 1)\cos(\tau - \eta)\big)\,. \end{aligned} \tag{3}$$

(a) Plot the potential for $2 \leq s \leq 5$, $1 \leq t \leq 2$ with $\eta = \tau = 0$, and for $-\pi \leq \eta, \tau \leq \pi$ with $s = 3$, $t = 1.1$ (why these values?). A contour plot is probably best-suited. Afterwards, narrow the boundaries such that you zoom in on the minimum.

(b) We want to use V as a fitness function, so it will be more natural to work with $-V$ from now on and maximize that. Note that $-V$ takes both positive and negative values. So either we have to shift it up by a constant or use a selection mechanism that can deal with negative fitness values. We will follow the latter route. Implement a binary tournament selection mechanism for mating.

(c) In order to trust our approximations we need $s, t > 1$. So we need a cross-over and mutation operation that respects this. For mutation, we will simply add a constant to any of the variables, drawn from a Gaussian with mean zero and variance 0.2. If this takes us to the regime $s, t < 1$, we will clip the value at 1. Implement this mutation routine.

(d) For crossover, we will use a version of `whole arithmetic crossover`, where we compute the children c_1 and c_2 via

$$c_1 = rx + (1 - r)y\,, \qquad c_2 = ry + (1 - r)x\,, \tag{4}$$

where r is drawn from a uniform distribution (between 0 and 1), and x, y are the parents, i.e. either of the s, t, η, τ. Implement the crossover routine.

(e) We use fitness-based survival and replace the 70 least fit parents with the 70 fittest offspring. Implement survival selection.

(f) Randomly initialize 100 individuals e_i (pay attention that $s, t \geq 1$). Each one will have four chromosomes, $e_i = (s_i, t_i, \eta_i, \tau_i)$. Run the algorithm for 300 generations with mutation rate 0.05 for each chromosome. Which value do you find? Compare with the results from (a). Why does the stabilization of s, t approximately decouple from the stabilization of η, τ for the parameter ranges considered here?