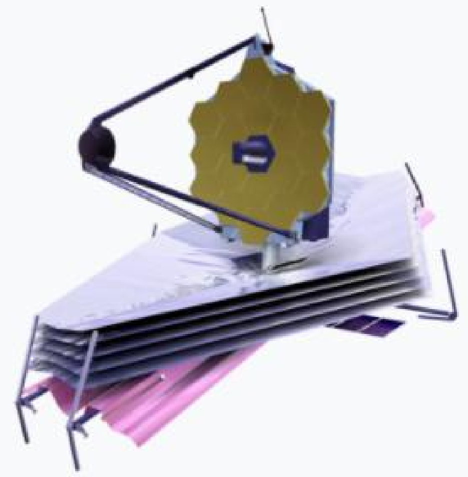
Planet Habitabilty using Machine Learning

With the developement of new advanced telescopes like the James Webb telescope, astronomy is turning into a big data problem.

The junction of Astronomy and Data science/ML is termed Astroinformatics

James Webb Space Telescope



A rendering of the James Webb Space Telescope with its components fully deployed.

Credits:Wikipedia

Our Goal:

To find out how suitable an exoplanet is for sustaining life using ML

We will use two approaches for this:

1. Fuzzy neural network classification
2. By finding a habitability score(CEESA score) for the exoplanet by optimizing a function Y

1.Fuzzy neural network classification

What are Fuzzy Sets?

https://www.youtube.com/watch?v=rln_kZbYaWc

Example of fuzzy set

$$X = \{5, 15, 25, 35, 45, 55, 65, 75, 85\}$$

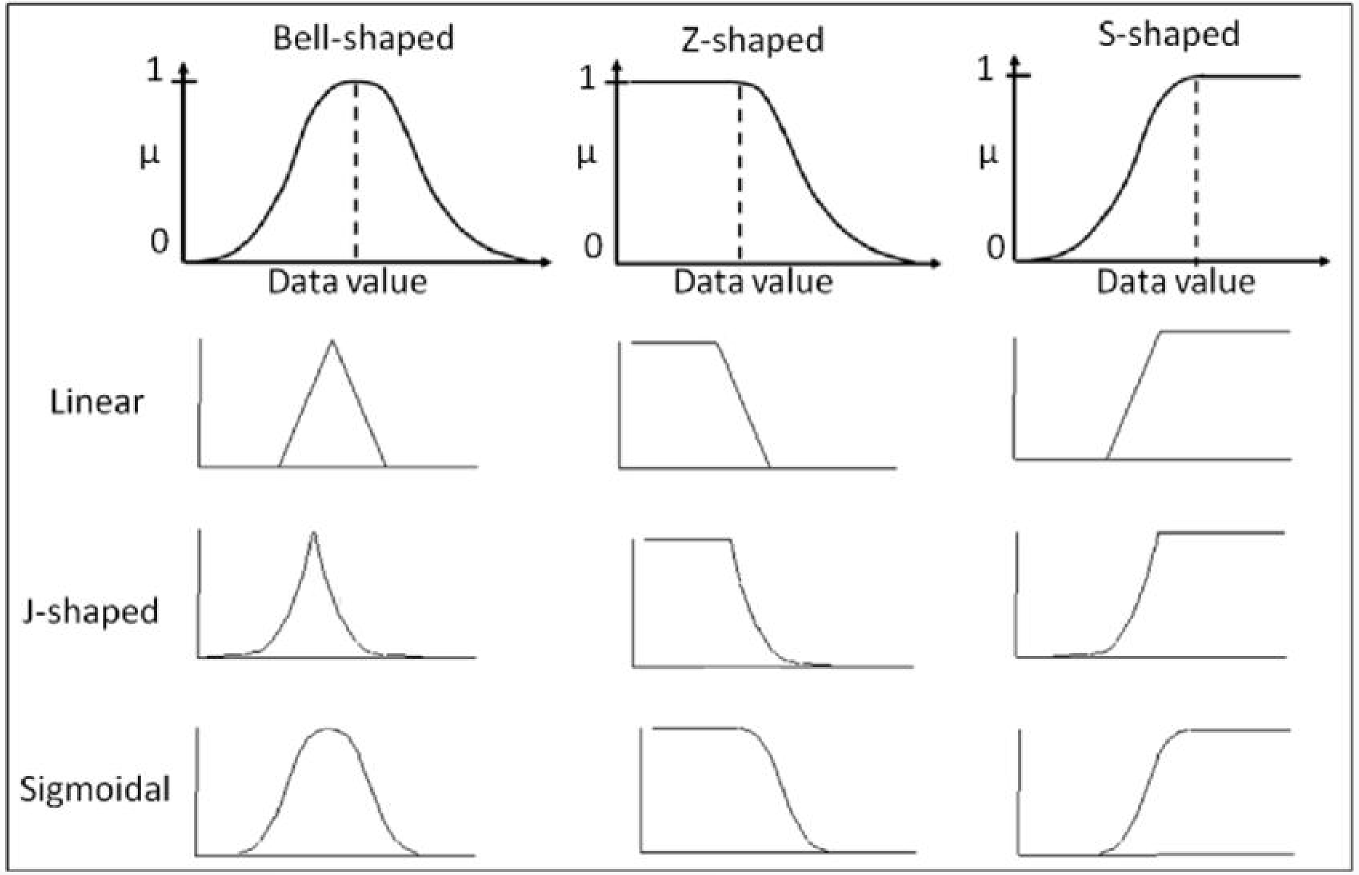
age(element)	infant	young	adult	senior
5	0	0	0	0
15	0	0.2	0.1	0
25	0	1	0.9	0
35	0	0.8	1	0
45	0	0.4	1	0.1
55	0	0.1	1	0.2
65	0	0	1	0.6
75	0	0	1	1
85	0	0	1	1

Membership values of element 35

How do we find these membership values?

Answer: Using Membership functions. They are functions which output the membership value y of an element x, given a fuzzy set A.

Membership function types

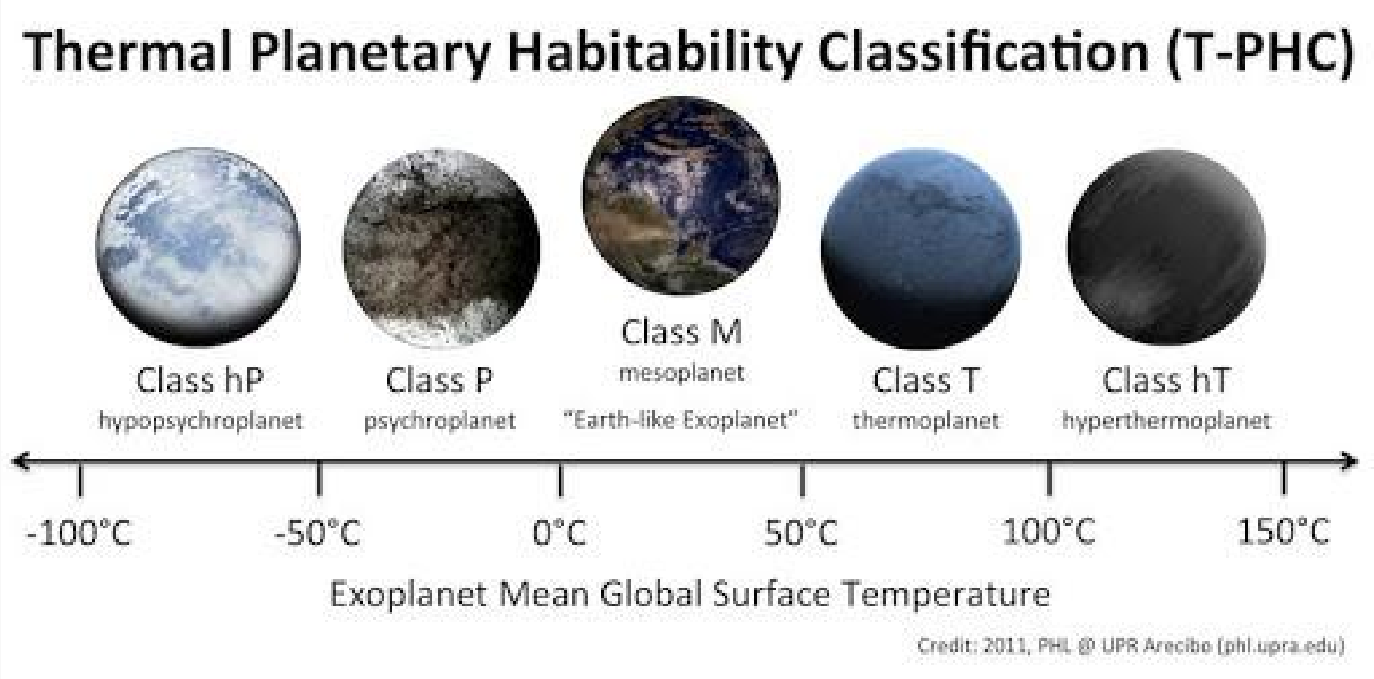


The ANN is trained on a total of 45 x 3 = 135 features and is used for classifying the test set into 3 classes(2 excluded):

Class 1 : non-habitable

Class 2 : mesoplanets(0 to 50° C)

Class 3 : psychroplanets(-50 to 0° C)



- **Pi membership function**. A membership function is an arbitrary curve that maps every value in the input space between 0 and 1. If X is the universe of discourse, x denotes an element, $\mu_A(x)$ is the membership function of x in A, then membership value is represented as $A = (\mu_A(x), x))$. The PI (π) function for a sample r (with c and λ as centre and radius of the dataset), can be defined as:

$$\pi(r; c, \lambda) = \begin{cases} 2(1 - \frac{\|r-c\|}{\lambda})^2 & \text{for} \quad \lambda/2 \le \|r - c\| \le \lambda \\ 1 - 2(\frac{\|r-c\|}{\lambda})^2 & \text{for} \quad 0 \le \|r - c\| \le \lambda/2 \\ 0 & \text{Otherwise} \end{cases}$$

Figure A.2 illustrates formation of 3 overlapping fuzzy sets using PI membership function.

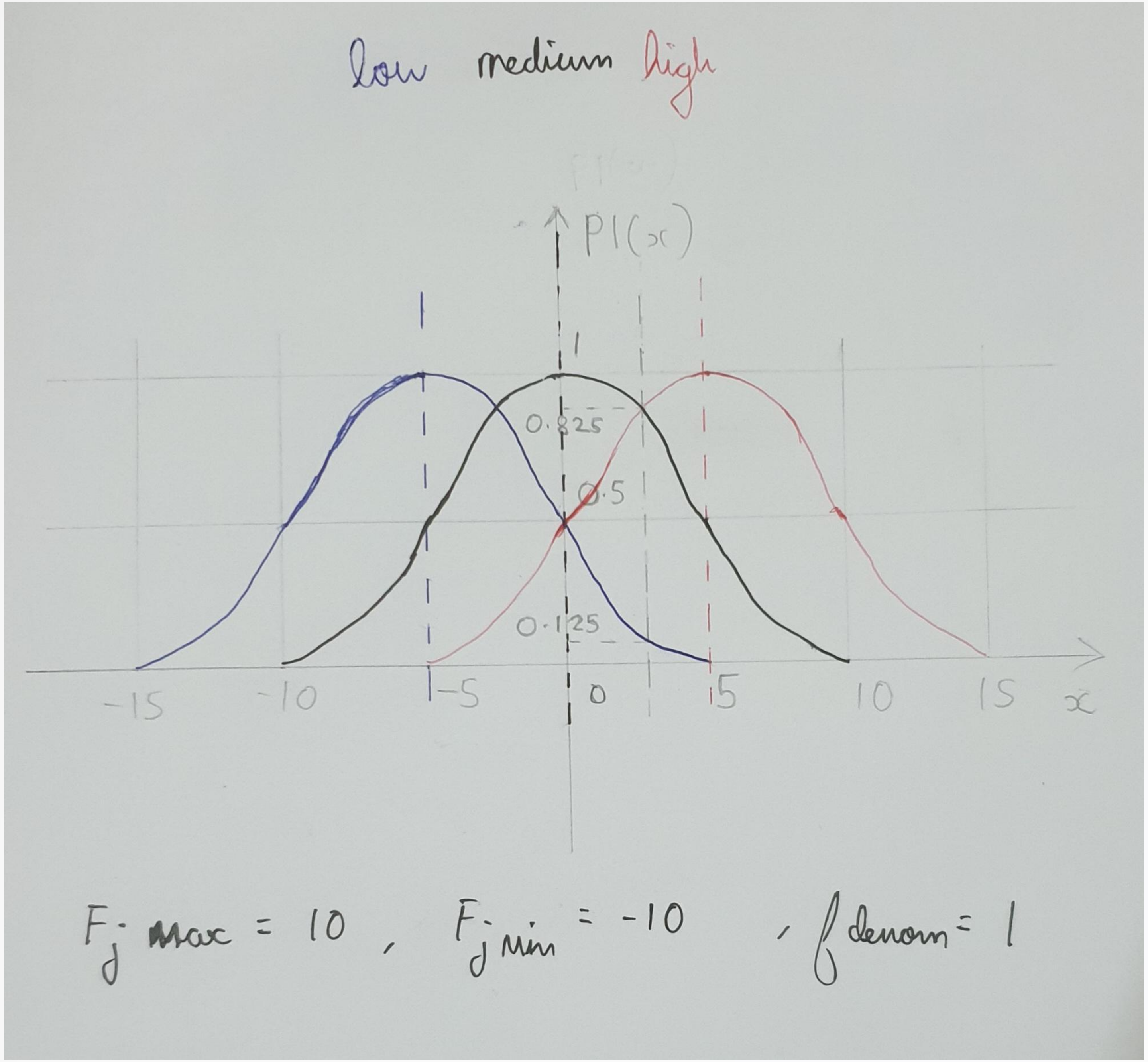
(parameter *fdenom* controls the level of overlap),

$$\lambda_{medium(Fj)} = \frac{1}{2}(F_{jMax} - F_{jMin})$$
$$c_{medium(Fj)} = F_{jMin} + \lambda_{medium(Fj)}$$
$$\lambda_{low(Fj)} = \frac{1}{fdenom}(c_{medium(Fj)} - F_{jMin})$$
$$c_{low(Fj)} = c_{medium(Fj)} - 0.5\,\lambda_{low(Fj)}$$
$$\lambda_{high(Fj)} = \frac{1}{fdenom}(F_{jMax} - c_{medium(Fj)})$$
$$c_{high(Fj)} = c_{medium(Fj)} + 0.5\,\lambda_{high(Fj)}\,.$$

For example, we take age from the earlier discussion.
Therefore, FjMax = 85 and FjMin = 5

We set fdenom=1
Then, on simplifying

$\lambda_{medium(F\,j)} = 40$
$c_{medium(F\,j)} = 45$
$\lambda_{low(F\,j)} = 40$
$c_{low(F\,j)} = 25$
$\lambda_{high(F\,j)} = 40$
$c_{high(F\,j)} = 65$



low medium high
PI(x)
1
0.825
0.5
0.125
-15
-10
-5
0
5
10
15
x
Fj max = 10, Fj min = -10, f denom = 1

2.By finding a habitability score(CEESA score)

A habitability score is found by maximizing a function Y.

$$Y = f(R, D, T_s, V_e, E) = (r.R^\rho + d.D^\rho + t.T_s^\rho + v.V_e^\rho + e.E^\rho)^{\frac{\eta}{\rho}} \ ,$$

where R is radius, D density, Ts surface temperature, Ve escape velocity and E the eccentricity of an exoplanet, which are given (in the dataset),

r, d, t, v, and e are the coefficients of radius, density, surface temperature, escape velocity and eccentricity, respectively. The coefficients lie in (0, 1) range,and Y is the target output. The sum of the coefficients r, d, t, v, and e should be 1.

The value of η is constrained by the scale of production used: $0 < \eta < 1$ under DRS, and $\eta = 1$ under CRS.

Y is the habitability score of exoplanets, where the aim is to maximize Y subject to the constraint that the range of ρ value is $0 < \rho \leq 1$.

We maximize Y using Particle Swarm Optimization(PSO).

Why not use the good old gradient ascent for this?

Ans:We can but it has less chance of finding the global maxima because of Curvature Violation

Curvature violation implies the change of sign in a functional form, i.e. the function changes its
shape (from increasing to decreasing, and vice-versa) prematurely even before the optima is reached.

PSO overcomes this problem and has a better chance of finding the global maxima.

PSO:

https://www.youtube.com/watch?v=JhgDMAm-imI&t=568s

After making some changes made to PSO implementaion in the paper to account for constraints, the algorithm looks like this:

Algorithm 1 Algorithm for CO by PSO.

Require: $f(x)$, the function to minimize.
Ensure: global minimum of $f(x)$.

```
for each particle i ← 1, n do
    repeat
        p_i ~ U(l, u)^d
    until p_i satisfies all constraints
    v_i ~ U(−|u − l|, |u − l|)^d
    pbest_i ← p_i
end for
gbest ← argmin_{pbest_i, i=1...n} f(pbest_i)
repeat
    oldbest ← gbest
    for each particle i ← 1 . . . n do
        u_p, u_g ~ U(0, 1)
        lbest ← argmin_{pbest_j, j=1...n} ||pbest_j − p_i||^2
        v_i ← ω.v_i + λ_g.u_g.(gbest − p_i) + λ_p.u_p.(lbest − p_i)
        p_i ← p_i + v_i
        if f(p_i) < f(pbest_i) and p_i satisfies all constraints then
            pbest_i ← p_i
        end if
    end for
    gbest ← argmin_{pbest_i, i=1...n} f(pbest_i)
until |oldbest − gbest| < threshold
return f(gbest)
```

Results: 1.Fuzzy ANN approach

Table 8: Case 2: Result of fuzzy classification of 3-class dataset with fuzzy inputs.

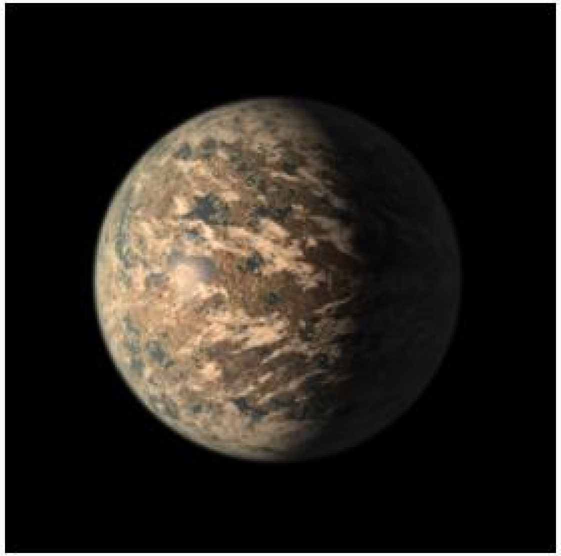
Class	Accuracy	Precision	Recall	Sensitivity	Specificity	Fscore
1	1.000	1.000	1.000	1.000	1.000	1.000
2	0.995	1.000	0.967	0.967	1.000	0.982
3	0.995	0.980	1.000	1.000	0.993	0.989

Class 1 : non-habitable
Class 2 : mesoplanets(0 to 50° C)
Class 3 : psychroplanets(-50 to 0° C)

2.Habitability score approach

Exoplanet	Habitability Score(CRS)	Habitability Score(DRS)
Earth	0.99	0.99
Kepler-186 f	1.15	0.99
Proxima Cen b	1.10	0.99
TRAPPIST-1 e	0.91	0.98
TRAPPIST-1 f	1.02	0.98
Ross 128 b	1.14	1.01

TRAPPIST-1e



Artist's impression of TRAPPIST-1e from 2018.

Credits:Wikipedia

The results were consistent with those of the Habitable Exoplanets Catalog(HEC)

http://phl.upr.edu/projects/habitable-exoplanets-catalog

So overall the validity of the approches was ascertained

The End

This presentation was based on this paper:

https://www.researchgate.net/publication/338314976_CEESA_meets_machine_learning_A_Constant_Elasticity_Earth_Similarity_Approach_to_habitability_and_classification_of_exoplanets