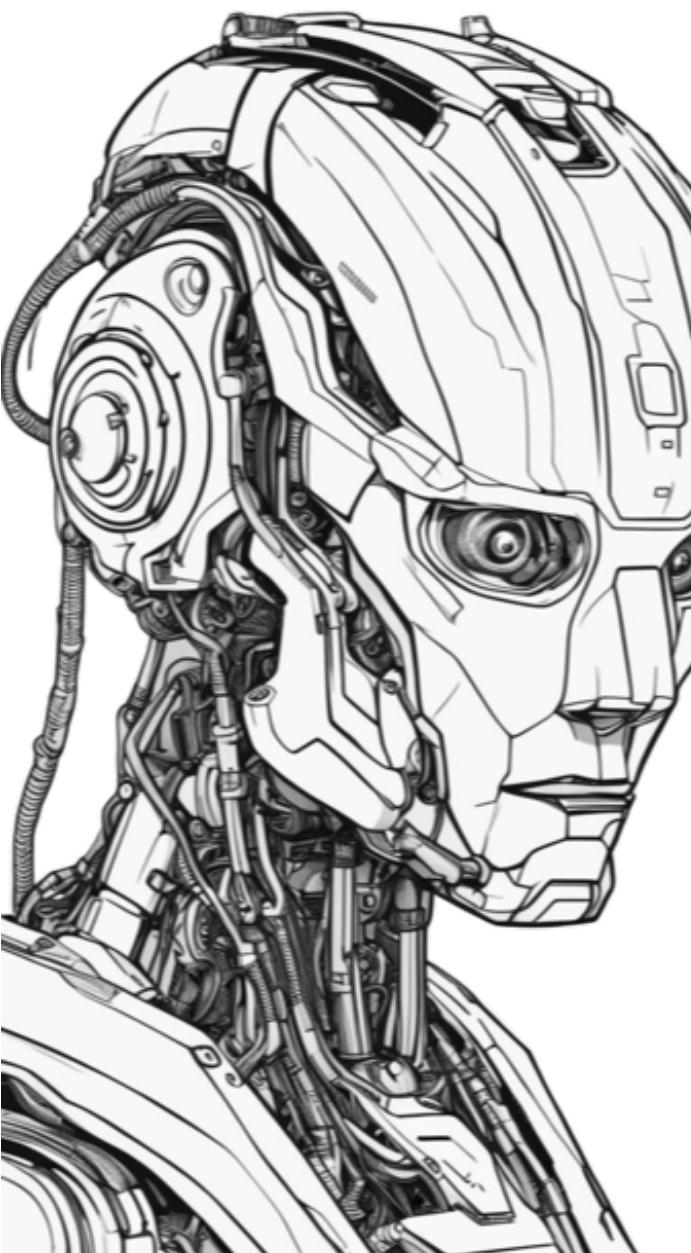


Artificial Intelligence Meets Biology

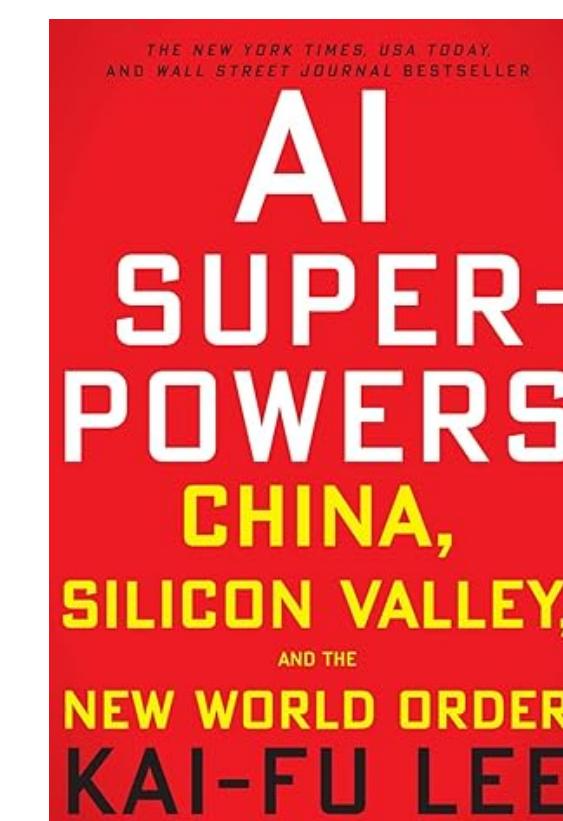


AI-based prediction of genomic elements.

Dr. Gustavo Sganzerla Martinez

“[...] with more data, an algorithm projected mid a mid-level team of engineers generally is better than one projected by an elite team of engineers.

Having the monopoly of the most brilliant minds does not mean having the best results.”



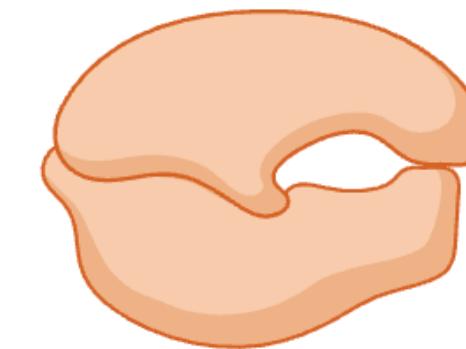
system



/'sistəm/

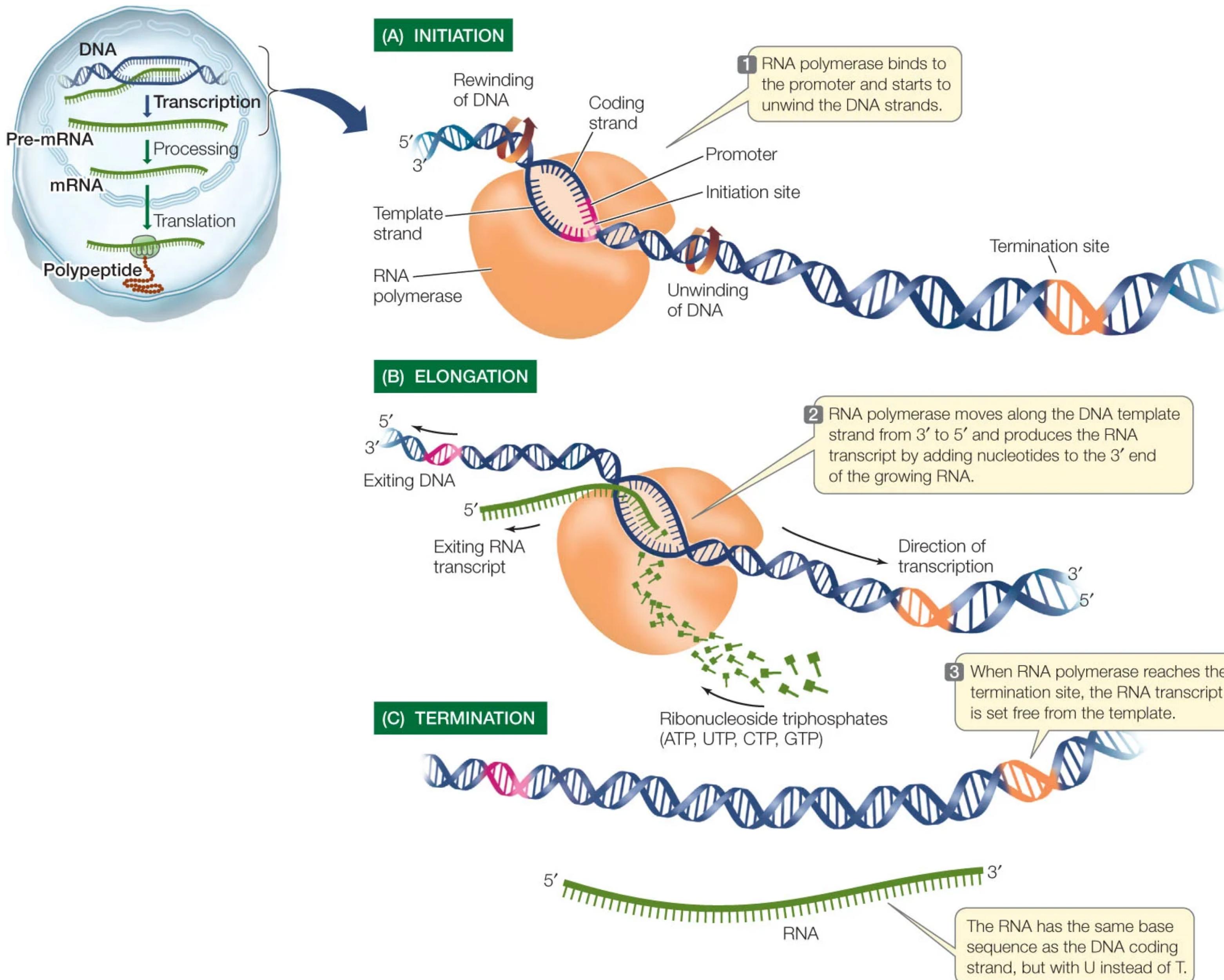
noun

- 1 a group or combination of interrelated, interdependent, or interacting elements forming a collective entity; a methodical or coordinated assemblage of parts, facts, concepts, etc



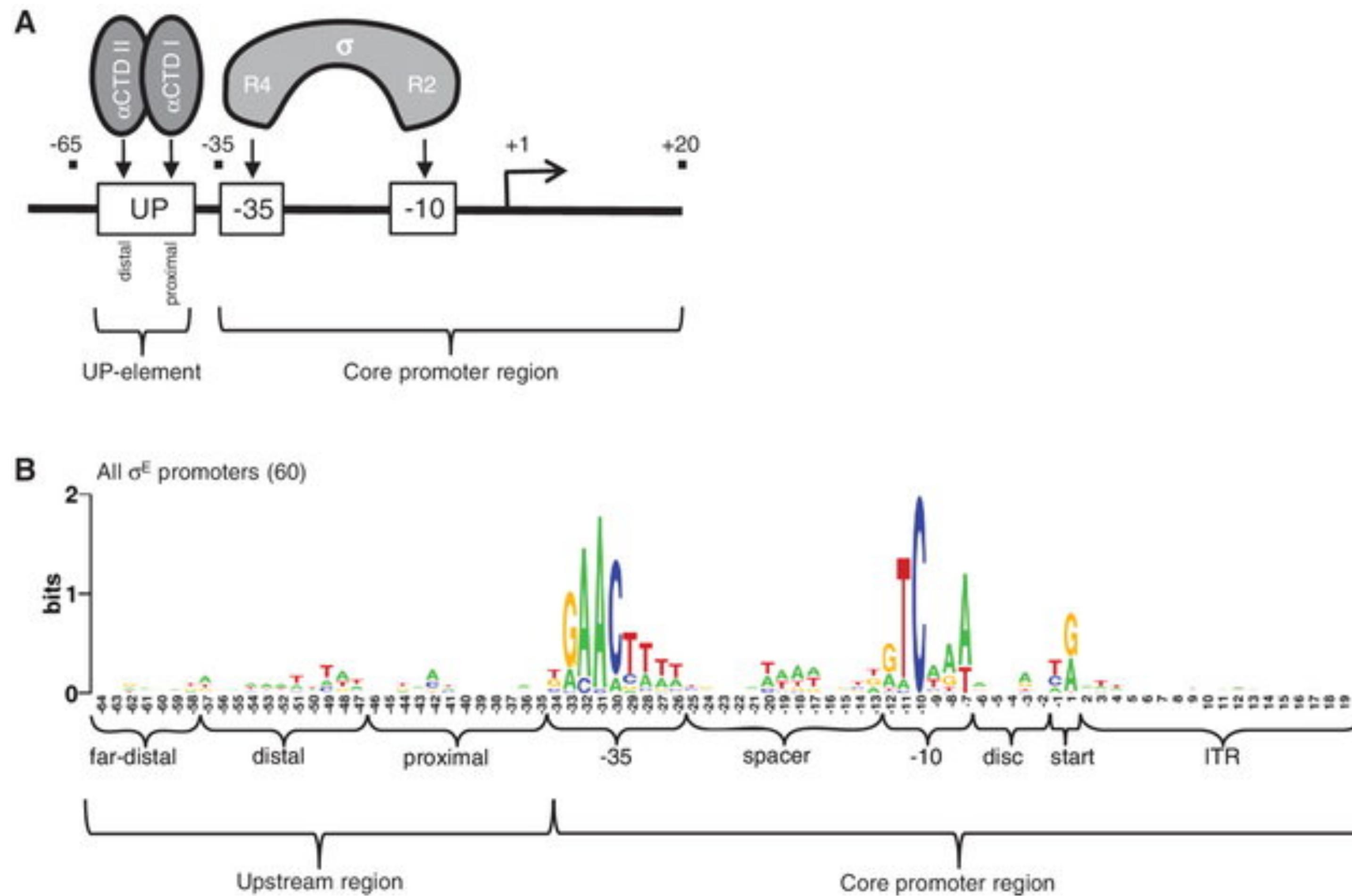
Input

Output



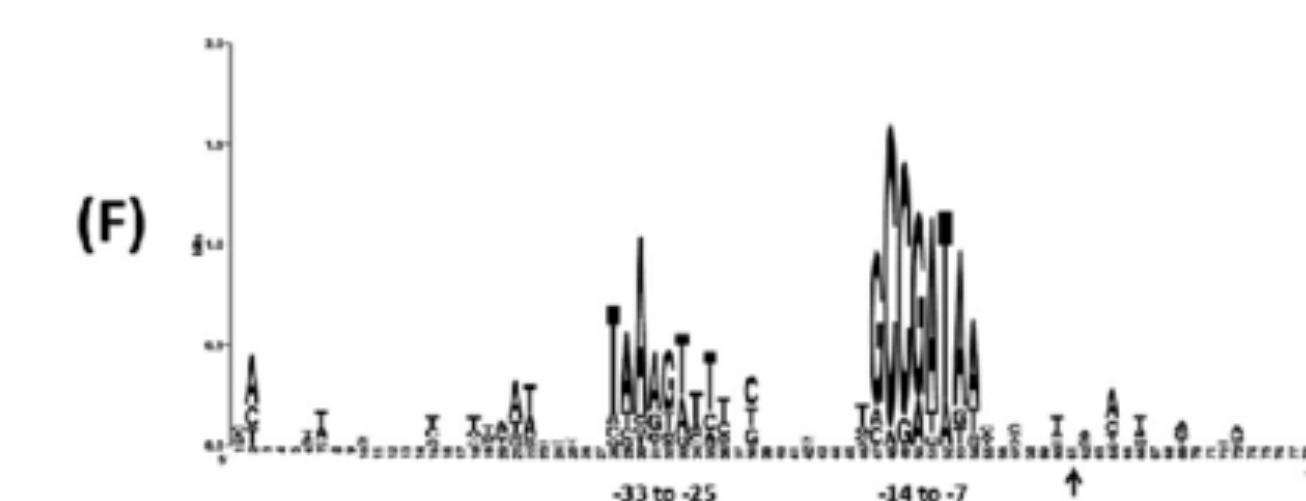
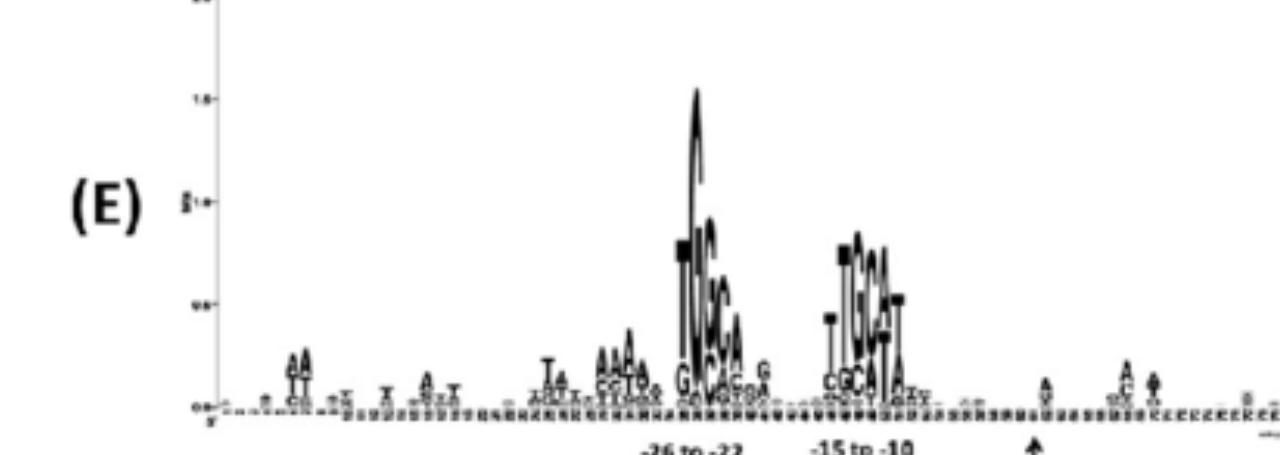
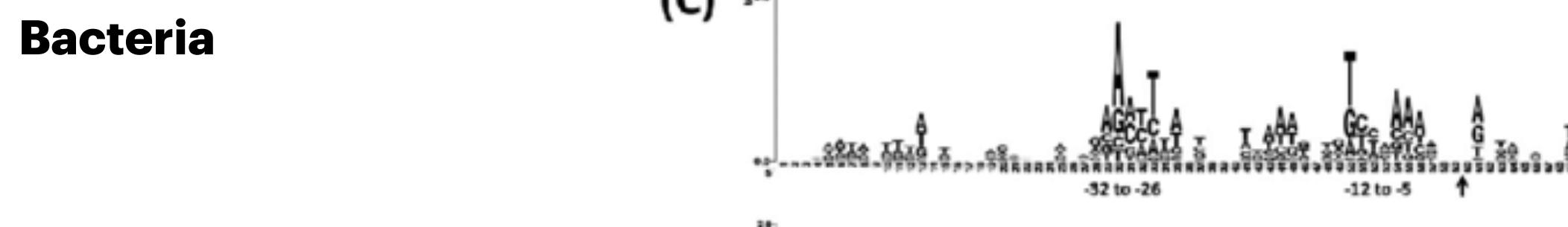
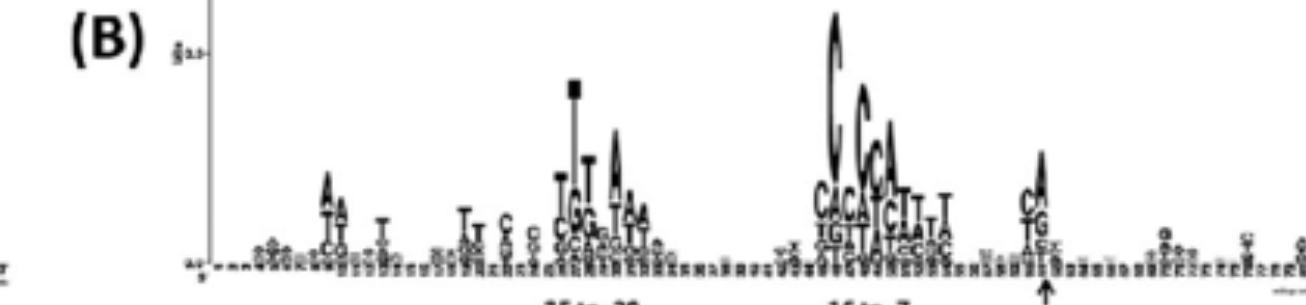
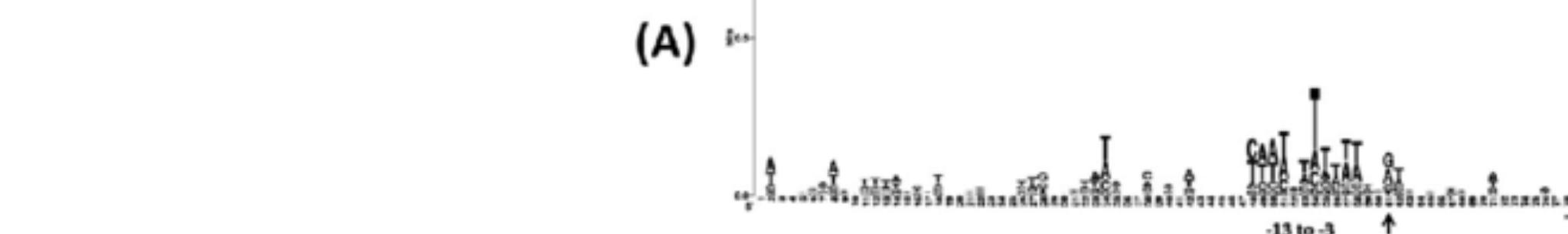
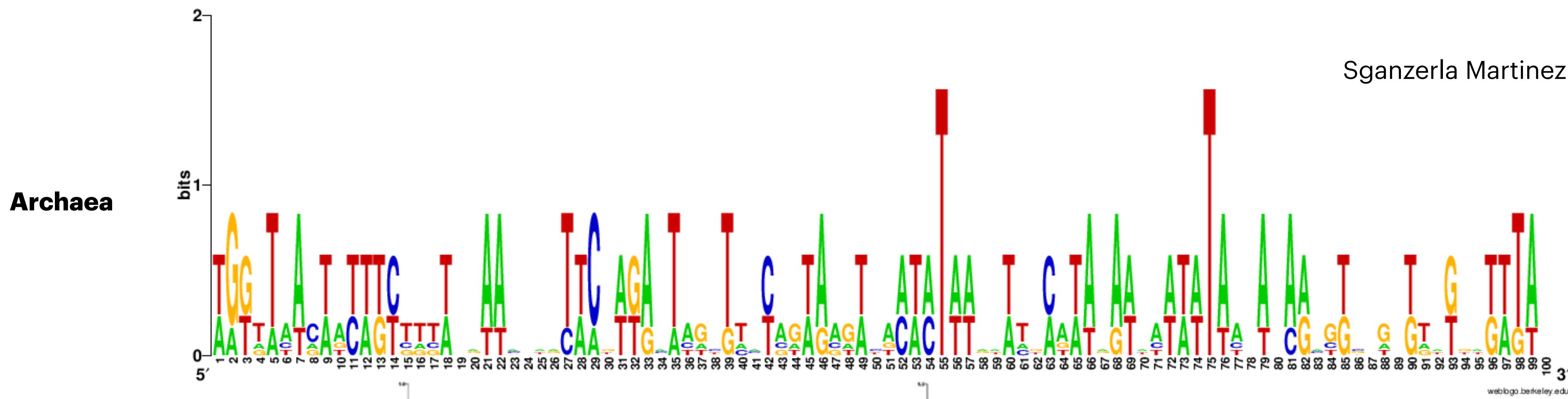
GTAGTAAATTATACATAATTTATAATTAATTACTATTTAGTGTCTAGAAAAAAATGTGTACCCACGACCGTAGGAAACTCTAGAGGGTAAGAAAATCAATCGTTAT
AGAGACCACAGAAAGAGGTTAATATTTGTGAGACCCATCGAAGAGAGAAAGGATAAAAACCTTTACGACTCCATCAGAAAGAGGTTAATATTTGTGAGACCCATCGACAGAG
AAAGAGATGGTTAGTCAAGATATTTCTTAGTACAAGAATGTTAAAATATGGACGAGAATTAATTGTCTGTATAAAAACCTGTGAAATTATGTACTAGAGAAAAACGTGA
GCAGTGTCCCCATGGATTACAGATCATTATTCACAAATATTAACATACGTTATTATGATGTTAACGTGTAACATTATAACATTATTTATGATGCAATTGCTGACAACCT
AGATTGGCATAAGGATATTGATAAGCTCTACGAGAATATATTGTTGGACGTTACGTTACGAAATAGTGAGACATCAGAAAGAGGTTAATATTTGTGAGACCATCGAAGAGAGAAAG
AGAATAAAAATATTTGTAAAACCTTTATGAGACAAGAGAGAAAGAGAATACGAATAGTGATCATCGTACATATTGAAACAGAAAGAAGTAACGAGAGGTAACCTTGT
GAATGTAGTTAACACATTGTTGCAAACCGGAATATAGTACCCGGTACACTTTAATTGTTGCGGTGTGAATCGTCGATTAACCCACTCATCCATTGAGATGAATAGA
GTTATCGATTGAGACACATGCTTGAGTTGAATCGATGAGTGAAGTATCATGGTGCACCTCAGATGCCGATCCGACACATGAAATCCATCCTGACTTCAAGTCAGAT
GATTCCCTCACACATGTCCTCGACGTACGCTAAACTCTAGGTTCTAACACATTGTATCAACGATCGTTGAACCGATGATATCTTGTAACACTACTTCTTATGTGAGATGTTAGACCC
AAGTACTGGATGGTCTTGATGTCGCTGTTCTCGTACATCTGATGTCGATAGACATCACAGTCTTGATCATGCCAGAGCTTACCGTACAGCGTACGCGGGAGAGTC
CTTACCTTGTCTGGTACACGCTGGACAATCTAGTATTCACAGTGTCCATCAGAGGATTGGAGATGGATGAAATCTTGGCATTGGTAATCCAAAGTTCATGTTAAGACCCG
CGCCGACGATAGTGAATAAGTGGTGGATCTCCTTACAACCTCTCGGATACCTCATCTCGGTCTGTAACTCCGTTACGGATTGACAAATCTTATCATGGTCTGGTTG
GTCTGCTTGTGACTTGATAATAACATCGATTCCCATATGATGTTCTTCTTCAGTACACGAGGATGAGGATTGTTGAAGACTAGTAGGCATAGCAGCTGCCACTAGGCACATG
CATGCCAGGACAATATATTGTTCATGATTGATTACTGTTCTAGATGATTCTACTTACCATATAAAATTAGAATATATTCTACTTACGAGAAATTAAATTGTATT
ATTATTATGGTAAAAAAACTACTATAAGTGGTGGATTCTGGATTAGTGATCAGTTATGTATATCGCAACTAGCAGGATATGGCTATTGACATCGAGAACATTACCCATATG
ATAAGAGATTGTATCCGTTCTGAGTATTGGTATTACTATAGTATGTAGATGTCGACGCTAGATAGACAGTCGCCCAGAGTTACCGTCTGAATGCCGATGATAGTAT
CATTCTTGTCTGTTACTGTTGGAAGATGAATCTTGTGTTACATTAACTCGAAATTAGTACAGAGTACACCGTTATTATACATGAGAGAAACAATATACGAGTATAACGGAC
CTCATGATTAATAATAGTAGTAATCGTCGTTACTGTTGTTCTACTCTCCAATCATATAGATTATTTAAATATTTCTTCTATCATGGATAATATTGTAATGGTCTTCCGTA
AACACATACTGTTAGATGGTAGTCGCTTAGCTGGTATGATATTGCGCATAATTCCGGAGGCAAATACGATAGTCTAGATTGACTATCGATGGTAGACTCTAATTGAGTGCTTGT
GACGAGTTACTTTATGCTCCATCGATAGATGACACTGTTCTATGAGATCGTACATGGAAATGAAATGCGTTGTGAATGTATGGCTCGAGATAGGTGTACCGGATGTCT
TCTGTTCTCAATACCGTACAGTTGGTCTGAGATTGAAATCTTGTGAGGAGACTTATGTCACGACTACATTTCGATGGAATCTATCGAATGATATTTTCAAATA
CACTTTATAGCCTCGTTAACAGAATTAGTAGTTCCGCAAATGACTCGCCCTTAATAGGCAGTAGGCTATTCTTACGTAGTGATCGTAGGGAGAGAAACTCCGC
ATCTTGTAGAACACGATTAATCATAGTAGAGATACTTCAGTCTGTGGATGATGTCATTACGACATCCGCCTGTATGTTCTGTTCAAACACCAAGTCGAATACTGT
CTTATCGTCTTAGTCGGAAGGTTGATGTCGTATCCGATGTATACGAGGTATGAGGCAACATTGTTATTGCAATTGGAAGGGCGGTATGAAGAGGAGTCATTGATTAGTATTGCTT
TCTGAATGTCGAATCTAGTACGAGAGAGCGACTCCATATCCTGATTATGTTATGAATAGATAAAAGTAGATGTTGTCCTTCTTGTAAATTCCCCTAT
TTTGTTCTGCCAATTGAGTAACATTGAGAATATGACCTGTTGCACAATCGTTCTTATGTATTCCATGATGGGTGACAATCAAGATTACGTATCCTCGATCGGCTCCTCGAGAA
AAAGAGCATAACACACACGAGGACTATGTTGGTACTGTTGAAGGTAAGTGTAAACGGCGTATTCCGATTTCGTAACCGCTTAATGTTGCTCCATGATCTATTACGCTAGAT
GAATCGCTCTCAGCTCGCATCTAGTGACTCTTGACTGTAATAATTGCTTCTGGAACACGGATATGTGTTACAGTAGTAATGAAGAGAAGTGAAGTCGACGCTACCCATCGACGCAA
TTAGGGTCAGATCCTTAGTCAATAATTGTACAGAACGTAATAGTTAACGCTCCATTGAATTAGTATTACTATAGTATGTAGATGTCGACGCTAGATAGACAGTCGCCCAGAGGTT
ACCGTCTCTGAATGCCGATGATAGTATCATTGTTCTGTTACTGTTGGAAGATGAATCTTGTGTTACATTAACTCGAAATTCAAGGTGCTGAGGATTGAAATCTTGTGAGGAGACTATGTCACG
AAATGCGTTGTCTGAGATGTATGGCTCGAGATAGGTGTACCGGATGTCTCTGTTCAATACCGTACAGTTGGTGTGAGGATTGAAATCTTGTGAGGAGACTATGTCACG
ACTACATTTCGATGATGGAATCTATCGAATGATATTTTCAAAATACACTTTATAGTCCTCGTTAACAGAATTAGTATGTAGTTCCGCAAATGACTCGCCCTTAATAGGC
AGTAGGCTATTATCTCTTACGTAGTGATCGTCGAGGGAGAGAACTCCGCATCTGTAGAACACGATTAACTCATAGTAGAGATACTTCAGTCTGTGGATGATGTCATTACG
ACATCCGCCTGTATGATGTTCAAACACCAAGTCGAATACTGTCTTACGTTAGTCGGAAGGTTGATGTCGATCGAGGTATGAGGCAACATTGTTATT

"Regulation of the gene expression ensures the right gene is expressed at the right moment."



Promoters are conserved DNA segments

Sganzerla Martinez et al, 2024. Unpublished



De Avila e Silva et al, 2013. *Biologicals*

GTAGTAAATTATACATAATTTATAATTAATTACTATTTAGTGTCTAGAAAAAAATGTGTACCCACGACCGTAGGAAACTCTAGAGGGTAAGAAAATCAATCGTTAT
AGAGACCACAGAAAGAGGTTAATATTTGTGAGACCCATCGAAGAGAGAAAGGATAAAAACCTTTACGACTCCATCAGAAAGAGGTTAATATTTGTGAGACCCATCGACAGAG
AAAGAGATGGTTAGTCAAGATATTTCTTAGTACAAGAATGTTAAAATATGGACGAGAATTAATTGTCTGTATAAAAACCTGTGAAATTATGTACTAGAGAAAAAACGTGA
GCAGTGTCCCCATGGATTACAGATCATTATTCACAAATATTAACATATACGTTATTATGATGTTACGTGAAATTATAAACATTATTTATGATGCAATTGCTGACAACCT
AGATTGGCATAAGGATATTGATAAGCTCTACGAGAATATATTGTTGGACGTTACGTTACGAAATAGTGAGACATCAGAAAGAGGTTAATATTTGTGAGACCATCGAAGAGAGAAAG
AGAATAAAAATATTTGTAAAACCTTTATGAGACAAGAGAGAAAGAGAATACGAATAGTGATCATATCGTATCACATATTGAAACAGAAAGAAGTAACGAGAGGTAACCTTGT
GAATGTAGTTAACACATTGTTGCAAACCGGAATATAGTACCCGGTACACTTTAATTGTTGCGGTGTGAATCGTCGATTAACCCACTCATCCATTGAGATGAATAGA
GTTATCGATTGAGACACATGCTTGAGTTGATGAGTGAAGTATCATGGTGCACCTCAGATGCCGATCCGACACATGAAATCCATCCTGACTTCAAGTCAGAT
GATTCCCTCACACATGTCCTCGACGTACGCTAAACTCTAGGTTCTAACACATTGTATCAACGATCGTTGAAACCGATGATATCTTGTAACTCACCTTCTATGTGAGATGTTAGACCC
AAGTACTGGATGGTCTTGATGTCGCTGTTCTCGCTACATCTGATGTCGATAGACATCACAGTCTTGATCATGCCAGAGCTTACCGTACAGCGTACGCGGGAGAGTC
CTTACCTTGTCTGGTACACGCTGGACAATCTAGTATTCACAGTGTCCATCAGAGGATTGGAGATGGATGAAATCTTGGCATTGGTAATCCAAAGTTCATGTTAAGACCCG
CGCCGACGATAGTGAATAAGTGGTGGATCTCCTTTACAACCTCTCGGATACCTCATCATCTCGGCTCTGTAACCTCCGTTACGGATTGACAAATCTTACATTGGTCTGGTTG
GTCTGCTTGTGACTTGATAATAACATCGATTCCCATATGATGTTCTTCTTCAGTACACGAGGATGAGGATTGTTGAAAGACTAGTAGGCATAGCAGCTGCCACTAGGCACATG
CATGCCAGGACAATATATTGTTCATGATTGATTACTGTTCTAGATGATTCTACTTACCATATAAAATTAGAATATATTCTACTTACGAGAAATTAAATTGTATT
ATTATTATGGTAAAAAAACTACTATAAGTGGTGGATTCTGGAAATTAGTGATCAGTTATGTATATCGCAACTAGCAGGATATGGCTATTGACATCGAGAACATTACCCATATG
ATAAGAGATTGTATCCGTTCTGAGTATTGGTATTACTATATAGTATGTAGATGTCGACGCTAGATAGACAGTCGCCCAGAGTTACCGTCTGAATGCCGATGATAGTAT
CATTCTTGTCTGTTACTGTTGGAAGATGAATCTGTTGTTACATTAACTCGAAATTAGTACACCCTTATTAT**TATA**CATGAGAGAAACAAT**TATA**CGAGTATAATACGGA
CCTCATGATTAATAATAGTAGTAATCGTCGTTACTGTTGTTCTACTCTCCAATCATATAGATTATTTAAATATTCTTCTATCATGGATAATATTGTAATGGTCTTCCG
TACAACATACTGTTAGATGGTAGTCGCTTAGCTGGTTATGATATTGCGCATAATTCCGGAGGCAAATACGATAGTCTAGATTGACTATCGATGGTAGACTCTAATTATTGAGTGCTT
TCGACGAGTTACTTTATGCTCCATCGATAGATGACACTGTTCTATGAGATCGTCGTACATGGAAATGCGTTGTGAATGTATGGCTCGAGATAGGTGATACCGGATGT
CTTCTGTTCTCAATACCGTATAAGTGGTCTGAGATTGAAATCTCTTGGAGACTATGTACGACTACATTTCGATGATGGAATCTATCGAATGATATTTTCAAA
TACACTTTATAGCCTCGTTAACAGAATTAGTATGTTCCGAAATGACTCGTCCCTTAATAGGCAGTAGGCTATTATCTTACGTAGTGTCTCGTAGGGAGAGAACTCCG
CATCTTGTAGAACACGATTAAATCATAGGTAGAGATACTTCAGTCTGTTGGATGATGTCTTAACGACATCCGCCCTGTATATGATGTTCTGTTCAAACACCAAGTCGAATACTG
TCTTATCGTCTTAGTCGAAGGTTGATCGTATCCGATGTATACGAGGTATGAGGCAACATTGTTATTGCAATTCTGGAAGGCAGGTATGAAGAGGAGTCATTGTATTAGTATTGCT
TTCTGAAATGTCGAATCTATCTAGATACCGTAGTATTGAGAGAGCGACTCCATATCCTGATTATGTTATGAAATAGATAAAAGTAGATGTTGTCCTCTTGTAAATTCCC
TTTTGTTGCGGCCATTGAGTAACATTGAGAATATGACCTGTTGCACAATCGTTCTTATGATTCCATGATGGGTACAATCAAGATTACGTATCCTCGTATCGGCTCCTCGAGA
AAAAGAGCATAACACCACACGAGGACTATGTTGGTACTGTTGAAGGTAAGTGTAAACGGCGTATTCCGATTTCGTAACCGCGTTAATGTTGCTCCATGATCTATTACGCT
TGAATCGCTCTCAGCTCGCATCTAGTGTGACTCTTGACTTGTAAATAATTGCTTCTGGAACACGGATATGTGTTACAGTAGTAATGAAGAGAAGTGAAGTCCATCCTCATCGACGCA
ATTAGGGTCAGATCCTTAGTCAATAATTGTACAGAACGTAATAGTTAACGCTCCATTGAATTAGTATTACTATATAGTATGTAGATGTCGACGCTAGATAGACAGTCGCCCAGAGT
TACCGTCTCTGAATGCCGATGATAGTATCATTCTGTTCTGTTAACTGTTGGAAGATGAATCTTGTGTTACATTAAATCTCGAAATTAGTACAGAGTATGAGATCGTCTGACATGGAAAT
GAAATCGTTGTCTGAATGTATGGCTCGAGATAGGTGTGATACCGGATGTCTCTGTTCAATACCGTATAAGTGGTCTGAGATTGAAATCTTGGAGAGACTATGTCAC
GACTACATTTCGATGATGGAATCTATCGAATGATATTTTCAAAATACACTTTATAGTCCTCGTTAAACAGAATTAGTATGTAGTTCCGAAATGACTCGTCCCTTAATAGG
CAGTAGGCTATTATCTCTTACGTAGTGTGACGAGAGAACTCCGCATCTGTAGAACACGATTAAATCATAGGTAGAGATACTTCAGTCTGTGGATGTCATTAAAC
GACATCCGCCATTGATGTTCTGTTCAAACACCAAGTCGAATACTGCTTACGTTAGTCGAAGGTTGATGTCGATGTACGAGGTATGAGGCAACATTGTTAT

Where does AI fit in?

What is AI?

IBM, 2024.

Artificial intelligence, or AI, is technology that enables computers and machines to simulate human intelligence and problem-solving capabilities.

Regression

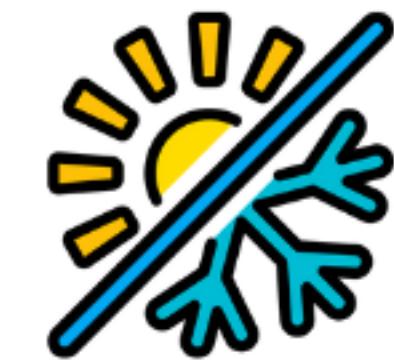


What will be the temperature tomorrow?

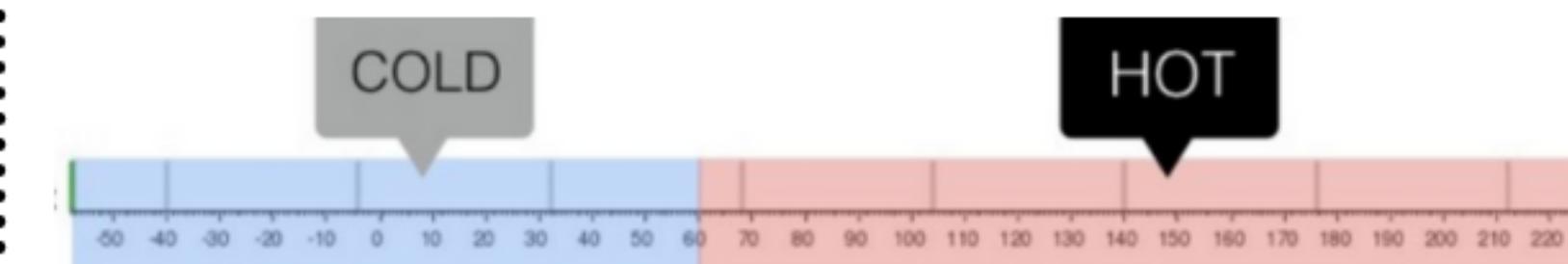


Fahrenheit

Classification

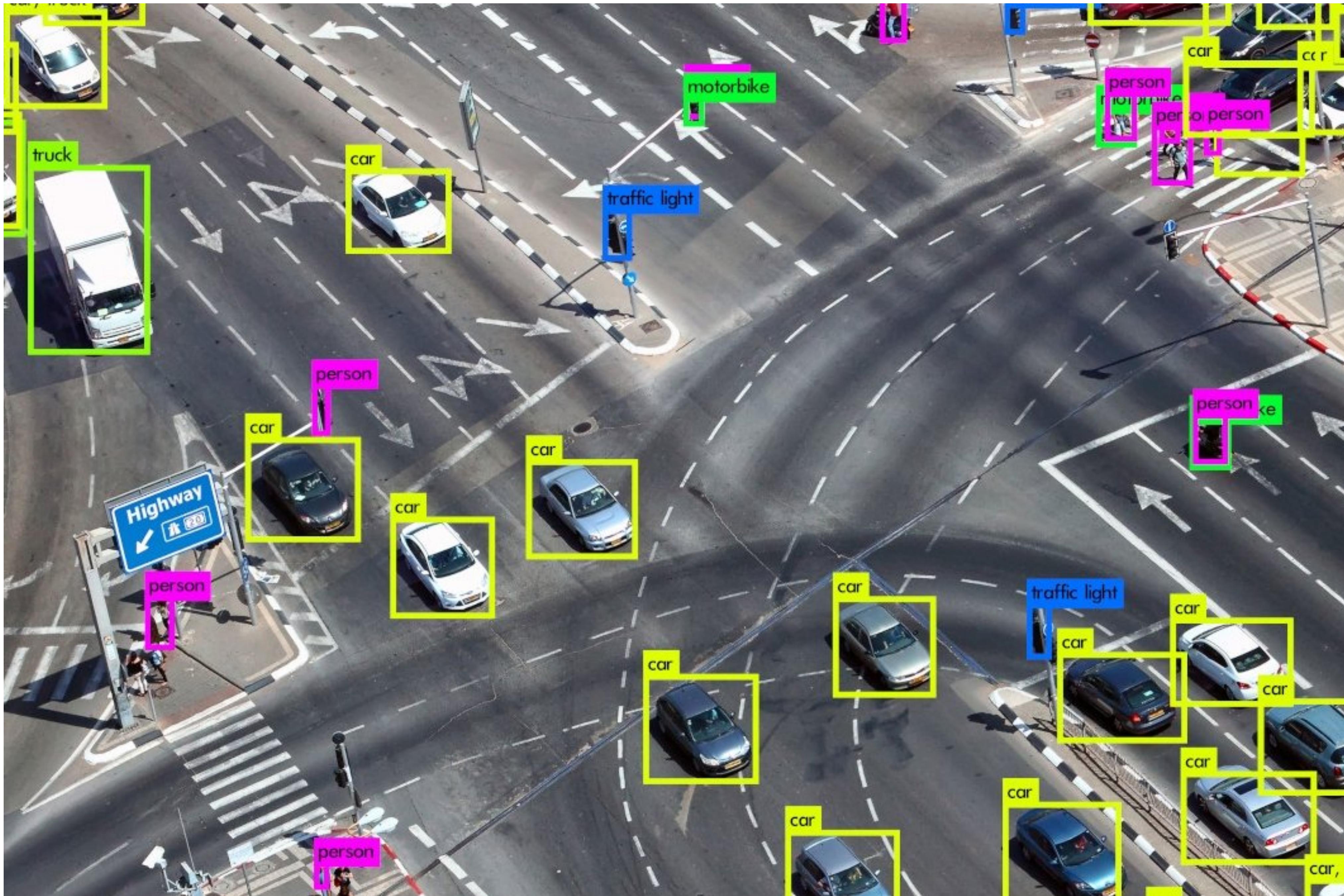


Will it be hot or cold tomorrow?



Fahrenheit

- Disease severity;
- Gene expression prediction;
- Population growth;
- Is a sequence a promoter?



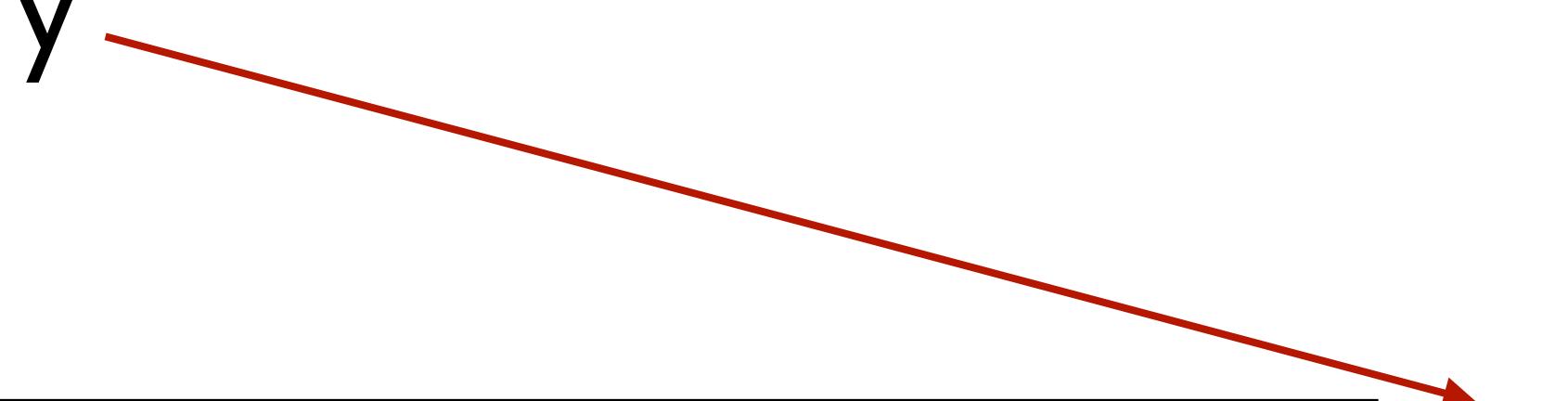
martinez.ace-net.training

usernames: user[001-099]

password: scarcely.daily.caring.redfish

X and y

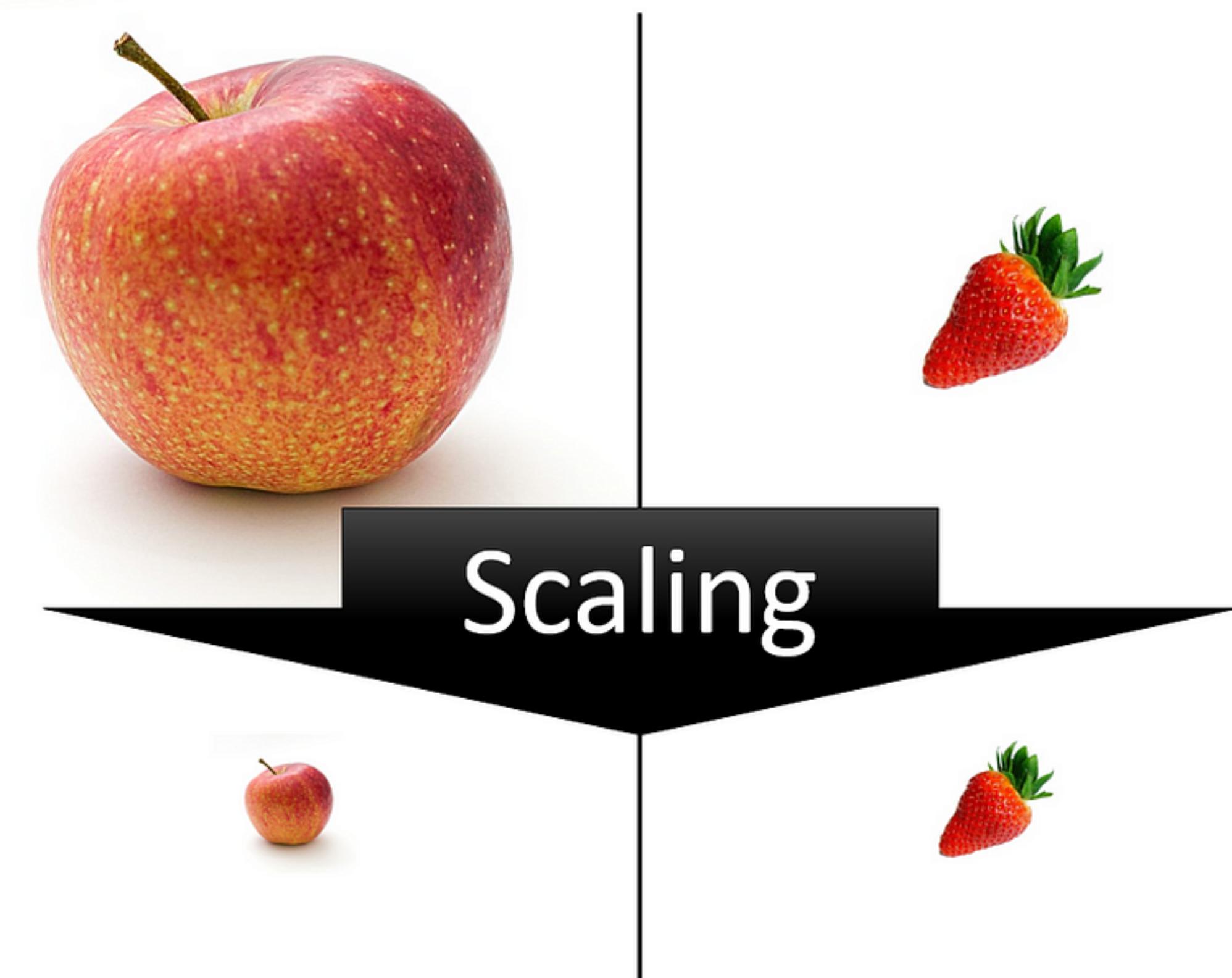
$$f(x) = y$$



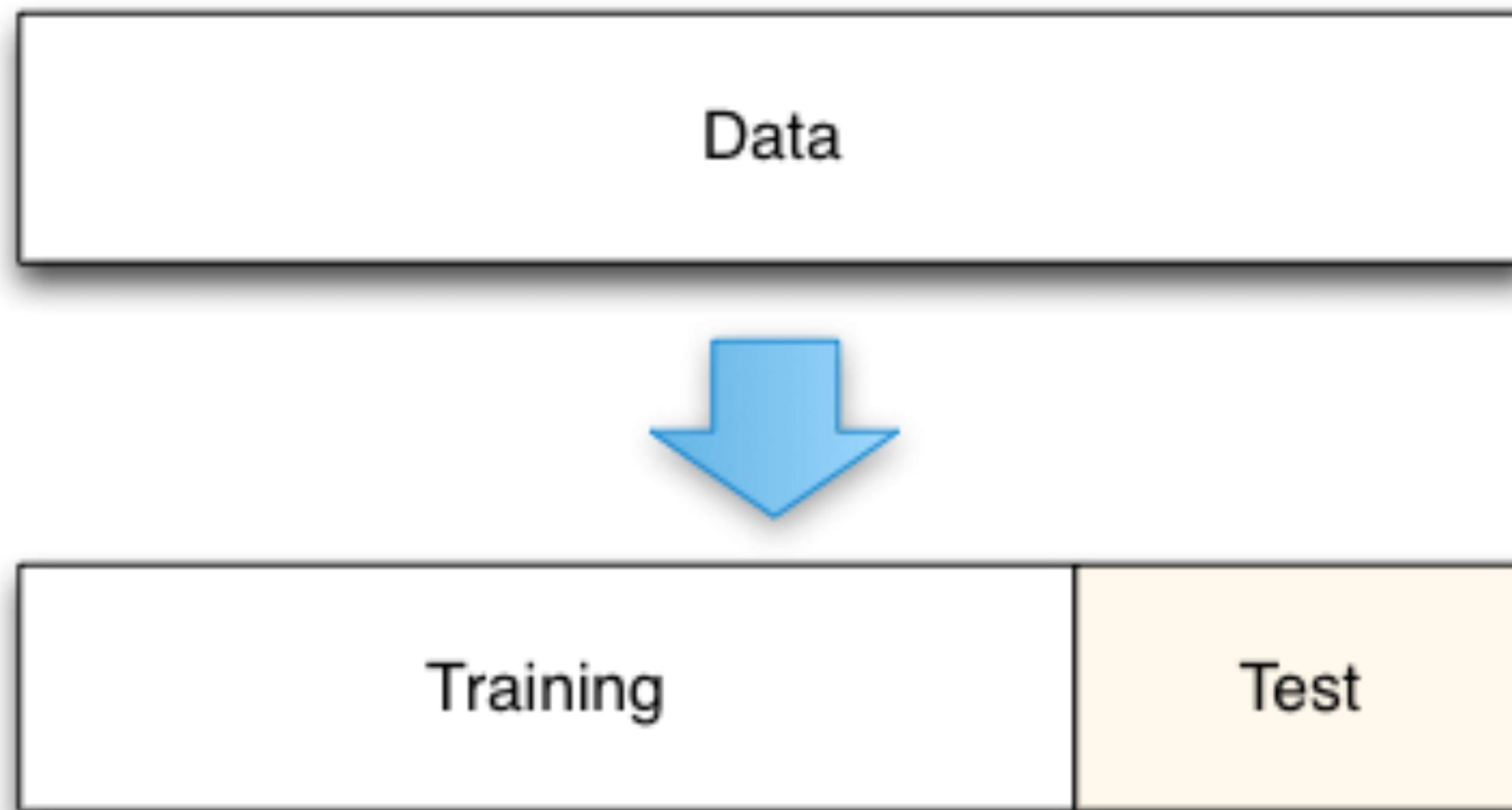
	0	1	2	3	4	5	6	7	8	9	...	50	51	52	53	54	55	56	57	58	label
0	-1.45	-1.30	-2.27	-1.28	-1.44	-2.27	-1.84	-2.24	-1.84	-1.44	...	-1.84	-1.30	-1.30	-1.44	-1.28	-1.28	-1.44	-1.30	-1.45	1
1	-1.28	-2.24	-1.30	-0.88	-1.28	-1.45	-1.00	-1.45	-1.28	-1.28	...	-1.30	-1.84	-1.30	-1.45	-1.28	-1.44	-1.84	-1.30	-0.88	1
2	-1.45	-1.45	-1.28	-0.58	-1.45	-1.45	-1.30	-1.44	-1.00	-0.58	...	-1.00	-1.44	-1.30	-1.30	-1.84	-1.30	-1.00	-0.88	-1.00	1
3	-0.88	-0.58	-0.88	-1.44	-1.84	-2.27	-1.84	-1.45	-1.45	-2.24	...	-2.27	-1.84	-1.45	-1.45	-1.45	-1.00	-1.00	-1.00	-1.00	1
4	-1.45	-1.28	-1.28	-1.84	-1.84	-1.45	-0.88	-0.58	-1.00	-1.30	...	-0.58	-1.00	-1.45	-2.24	-1.44	-1.44	-2.27	-1.84	-2.24	1
...	
3645	-0.88	-0.58	-1.30	-1.44	-0.58	-0.88	-1.28	-2.24	-2.27	-1.28	...	-1.00	-0.58	-0.88	-0.58	-0.88	-0.58	-1.00	-1.00	-1.00	0
3646	-1.00	-0.88	-1.00	-1.00	-0.58	-1.45	-1.84	-1.28	-0.58	-1.45	...	-1.84	-1.45	-1.30	-1.30	-0.88	-1.00	-1.28	-1.84	-1.45	0
3647	-1.30	-1.30	-1.30	-1.45	-1.28	-0.58	-1.30	-1.30	-0.88	-1.00	...	-0.88	-1.00	-1.44	-1.44	-1.00	-1.00	-0.58	-1.45	-2.24	0
3648	-1.28	-1.28	-1.45	-0.88	-1.00	-1.00	-1.44	-1.84	-1.30	-0.88	...	-0.88	-1.28	-2.24	-1.44	-1.28	-1.84	-2.24	-1.44	-1.00	0
3649	-1.00	-1.30	-1.30	-1.00	-0.88	-1.28	-1.28	-1.00	-0.58	-0.88	...	-1.44	-1.00	-1.44	-1.44	-1.00	-0.58	-0.88	-0.58	-1.45	0

3650 rows × 60 columns

Data scaling



Data split



Model interpretation

