Antigen Classification Problem

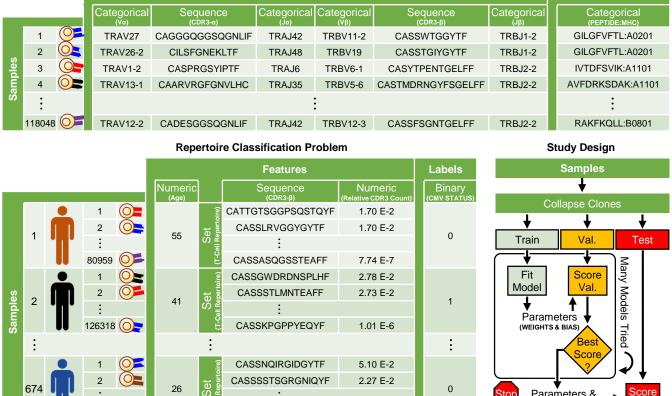|	Features						Labels
	Categorical (Vα)	Sequence (CDR3-α)	Categorical (Jα)	Categorical (Vβ)	Sequence (CDR3-β)	Categorical (Jβ)	Categorical (PEPTIDE:MHC)
1	TRAV27	CAGGGQGGSQGNLIF	TRAJ42	TRBV11-2	CASSWTGGYTF	TRBJ1-2	GILGFVFTL:A0201
2	TRAV26-2	CILSFGNEKLTF	TRAJ48	TRBV19	CASSTGIYGYTF	TRBJ1-2	GILGFVFTL:A0201
3	TRAV1-2	CASPRGSYIPTF	TRAJ6	TRBV6-1	CASYTPENTGELFF	TRBJ2-2	IVTDFSVIK:A1101
4	TRAV13-1	CAARVRGFGNVLHC	TRAJ35	TRBV5-6	CASTMDRNGYFSGELFF	TRBJ2-2	AVFDRKSDAK:A1101
⋮							⋮
118048	TRAV12-2	CADESGGSQGNLIF	TRAJ42	TRBV12-3	CASSFSGNTGELFF	TRBJ2-2	RAKFKQLL:B0801

Samples

Repertoire Classification Problem

		Features		Labels
	Numeric (Age)	Sequence (CDR3-β)	Numeric (Relative CDR3 Count)	Binary (CMV STATUS)
1	55	CATTGTSGGPSQSTQYF	1.70 E-2	0
		CASSLRVGGYGYTF	1.70 E-2	
		⋮		
		CASSASQGSSTEAFF	7.74 E-7	
2	41	CASSGWDRDNSPLHF	2.78 E-2	1
		CASSSTLMNTEAFF	2.73 E-2	
		⋮		
		CASSKPGPPYEQYF	1.01 E-6	
⋮				
674	26	CASSNQIRGIDGYTF	5.10 E-2	0
		CASSSSTSGRGNIQYF	2.27 E-2	
		⋮		
		CASSLGRLSSYEQYF	2.81 E-6	

Samples

Set (T-Cell Repertoire)

Study Design

Samples → Collapse Clones → Train / Val. / Test

Train → Fit Model → Parameters (WEIGHTS & BIAS)

Val. → Score Val. → Best Score?

Many Models Tried

Best Score? → Stop → Parameters & Hyperparameters (WEIGHTS, BIAS, & CUTOFF) → Score Test

Test → Score Test