Original Training Dataset [Used to Train Base Estimators]

	pixel1	pixel2	pixel3	pixel4	● ● ●	pixel784	Target
d_0	123	124	137	141		0	shirt
d_1	11	12	10	09		216	sneakers
				● ● ●			
d_n	241	238	251	254		248	trousers

28 * 28 = 784 pixels gray level values per instance plus a target feature value

Stack Layer Training Set (Labels) [Task 1]

	$\mathbb{M}_0$	$\mathbb{M}_1$	$\mathbb{M}_2$	$\mathbb{M}_3$	•••	$\mathbb{M}_e$	Target
d_0	shirt	shirt	shirt	sneakers		shirt	shirt
d_1	trousers	shirt	trousers	sneakers		shirt	sneakers
				⋮			
d_n	shirt	trousers	sneakers	shirt		trousers	trousers

Each training instance is composed of the class prediction by each base estimator plus the target feature value for an instance from the original training dataset

Stack Layer Training Set (Probabilities) [Task 3]

$\mathbb{M}_0$			$\mathbb{M}_1$			$\mathbb{M}_2$			$\mathbb{M}_3$			●●● $\mathbb{M}_e$			Target
d_0 0.81	0.11	0.08	0.76	0.14	0.10	0.99	0.01	0.00	0.32	0.65	0.03	0.91	0.06	0.02	shirt
d_1 0.38	0.07	0.55	0.78	0.13	0.09	0.17	0.23	0.60	0.10	0.58	0.32	0.76	0.16	0.08	sneakers
d_n 0.92	0.04	0.04	0.05	0.06	0.89	0.02	0.76	0.22	0.54	0.02	0.44	0.06	0.11	0.83	trousers

shirt	sneakers	trousers

* Assuming just three possible target classes: {shirt, sneakers, trousers}

Modified Stack Layer Training Set (Probabilities + Original Features) [Task 8]

	pixel 1	pixel 2	pixel 3	•••	pixel 784	$\mathbb{M}_0$			$\mathbb{M}_1$			•••	$\mathbb{M}_e$			Target
d_0	123	124	137		0	0.81	0.11	0.08	0.76	0.14	0.10		0.91	0.06	0.02	shirt
d_1	11	12	10		216	0.38	0.07	0.55	0.78	0.13	0.09		0.76	0.16	0.08	sneakers
		• • •						• • •						• • •		
d_n	241	238	251		248	0.92	0.04	0.04	0.05	0.06	0.89		0.06	0.11	0.83	trousers

Add the original training instance to the start of the dataset used to train the stacked layer - so original training features plus base estimator predictions