



# Correlação em Machine Learning

# Aula	42
<input checked="" type="checkbox"/> Preparada	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Revisada	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Lecionada	<input checked="" type="checkbox"/>

## ▼ Analisar a Correlação é extremamente útil em Machine Learning

- Variáveis preditoras muito correlacionadas podem interferir na performance de alguns modelos de Machine Learning...
- Ao mesmo tempo é bom ter variáveis preditoras bem correlacionadas com a variável resposta...
- ...mesmo que a relação entre variáveis preditoras e a variável resposta não seja linear.

Ainda assim, podemos medir a correlação apenas entre 2 variáveis quantitativas.

## ▼ Surge a ideia da Matriz de Correlação

Nesta matriz temos as variáveis nas linhas e repetidas nas colunas...

## ▼ Normalmente expressa no já mencionado Mapa de Calor:

	Var_01	Var_02	Var_03	Var_04	Var_05	Var_06	Var_07	Var_08	Var_09	Var_10	Var_11	Var_12	Var_13	Var_14	Var_15	Var_16	Var_17	Var_18	Var_19	Var_20
Var_01	1,00	0,52	0,21	-0,87	-0,36	-0,32	-0,73	-0,85	-0,20	0,67	0,72	-0,50	0,14	-0,67	0,34	0,77	0,50	0,48	-0,34	0,05
Var_02	0,52	1,00	0,60	0,13	0,52	-0,69	0,63	0,38	0,15	0,15	0,39	0,62	0,50	0,90	-0,38	-0,30	-0,72	0,00	-0,52	0,23
Var_03	0,21	0,60	1,00	-0,74	0,35	-0,25	-0,22	0,82	-0,51	0,95	0,71	0,98	-0,29	0,91	0,94	0,07	0,59	-0,78	-0,08	0,08
Var_04	-0,87	0,13	-0,74	1,00	0,81	-0,07	-0,71	0,21	-0,40	-0,24	0,69	0,45	0,47	0,08	0,69	-0,03	0,63	-0,71	-0,76	0,64
Var_05	-0,36	0,52	0,35	0,81	1,00	0,62	-0,74	-0,83	0,30	0,97	0,53	-0,63	-0,64	0,45	0,39	0,30	-0,18	0,78	0,02	0,68
Var_06	-0,32	-0,69	-0,25	-0,07	0,62	1,00	0,99	-0,51	0,27	-0,12	-0,04	-0,84	0,14	-0,32	0,73	0,07	0,33	-0,90	0,62	-0,32
Var_07	-0,73	0,63	-0,22	-0,71	-0,74	0,99	1,00	0,36	-0,69	-0,45	0,68	-0,01	-0,89	0,20	-0,83	0,54	-0,06	0,28	0,07	0,98
Var_08	-0,85	0,38	0,82	0,21	-0,83	-0,51	0,36	1,00	0,47	0,21	0,32	-0,05	0,96	-0,71	-0,95	0,53	-0,26	-0,54	-0,03	0,99
Var_09	-0,20	0,15	-0,51	-0,40	0,30	0,27	-0,69	0,47	1,00	-0,64	-0,75	0,20	0,24	-0,51	-0,69	-0,11	-0,32	0,36	-0,32	-0,19
Var_10	0,67	0,15	0,95	-0,24	0,97	-0,12	-0,45	0,21	-0,64	1,00	-0,12	0,11	-0,72	0,79	0,30	0,98	0,60	0,67	-0,50	-0,89
Var_11	0,72	0,39	0,71	0,69	0,53	-0,04	0,68	0,32	-0,75	-0,12	1,00	0,25	-0,95	-0,88	-0,93	0,73	0,54	0,20	-0,18	0,31
Var_12	-0,50	0,62	0,98	0,45	-0,63	-0,84	-0,01	-0,05	0,20	0,11	0,25	1,00	0,68	-0,39	0,43	0,29	-0,42	0,63	0,51	-0,41
Var_13	0,14	0,50	-0,29	0,47	-0,64	0,14	-0,89	0,96	0,24	-0,72	-0,95	0,68	1,00	0,17	0,14	0,29	-0,62	0,90	-0,67	-0,22
Var_14	-0,67	0,90	0,91	0,08	0,45	-0,32	0,20	-0,71	-0,51	0,79	-0,88	-0,39	0,17	1,00	0,10	-0,90	0,11	0,26	0,19	0,40
Var_15	0,34	-0,38	0,94	0,69	0,39	0,73	-0,83	-0,95	-0,69	0,30	-0,93	0,43	0,14	0,10	1,00	-0,87	-0,20	-0,39	-0,57	-0,50
Var_16	0,77	-0,30	0,07	-0,03	0,30	0,07	0,54	0,53	-0,11	0,98	0,73	0,29	0,29	-0,90	-0,87	1,00	0,68	-0,29	-0,17	0,58
Var_17	0,50	-0,72	0,59	0,63	-0,18	0,33	-0,06	-0,26	-0,32	0,60	0,54	-0,42	-0,62	0,11	-0,20	0,68	1,00	0,53	-0,14	-0,54
Var_18	0,48	0,00	-0,78	-0,71	0,78	-0,90	0,28	-0,54	0,36	0,67	0,20	0,63	0,90	0,26	-0,39	-0,29	0,53	1,00	-0,93	-0,37
Var_19	-0,34	-0,52	-0,08	-0,76	0,02	0,62	0,07	-0,03	-0,32	-0,50	-0,18	0,51	-0,67	0,19	-0,57	-0,17	-0,14	-0,93	1,00	-0,18
Var_20	0,05	0,23	0,08	0,64	0,68	-0,32	0,98	0,99	-0,19	-0,89	0,31	-0,41	-0,22	0,40	-0,50	0,58	-0,54	-0,37	-0,18	1,00

## ▼ Propriedades da Matriz de Correlação

Como  $\text{Correlação}(x, x) = 1$ , a diagonal da matriz de correlação é sempre igual a 1.

Como  $\text{Correlação}(x, y) = \text{Correlação}(y, x)$ , a matriz de correlação é sempre simétrica.