



# DÉFICIT HABITACIONAL

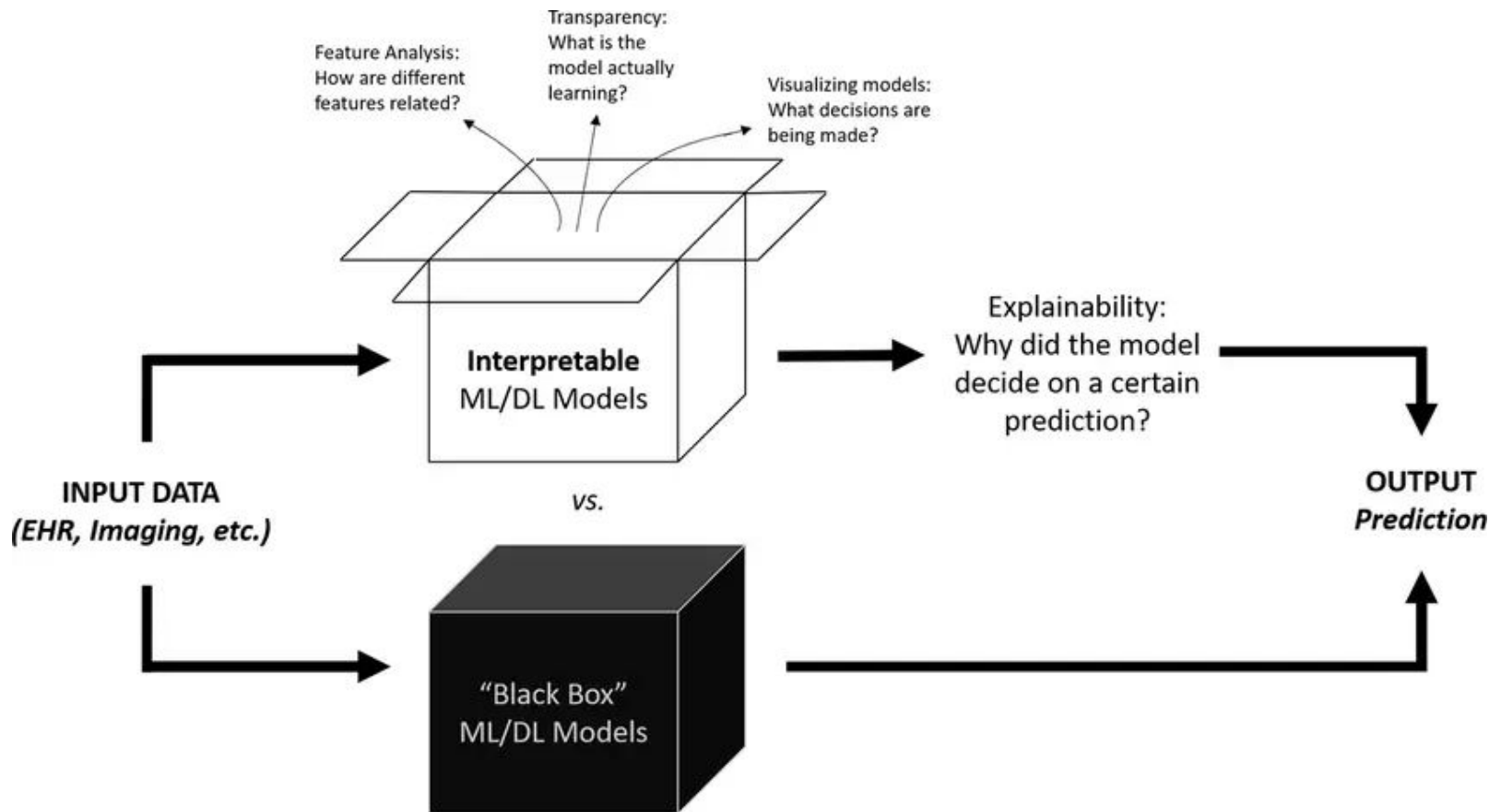
MINERAÇÃO DE DADOS, PADRÕES DE APRENDIZADO, APRENDIZADO SUPERVISIONADO,  
CONCEITOS DE ENTRADA E TIPOS DE APRENDIZADO

# 1. SOBRE A MINERAÇÃO DE DADOS E O APRENDIZADO DE MÁQUINAS

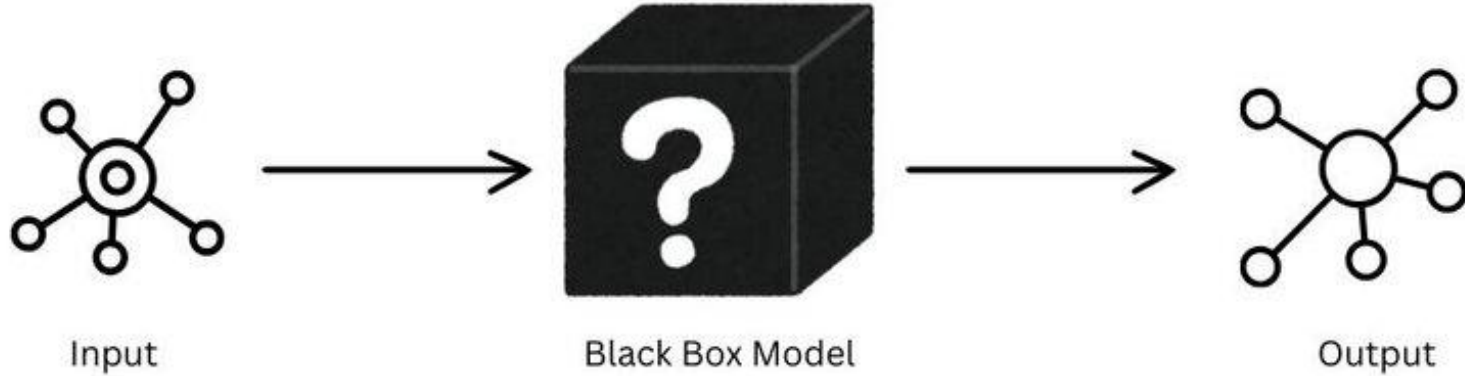
ÁREAS DA COMPUTAÇÃO

---

# PADRÕES DE MODELOS DE APRENDIZADO



# PADRÕES DE MODELOS DE APRENDIZADO



# PADRÕES DE MODELOS DE APRENDIZADO

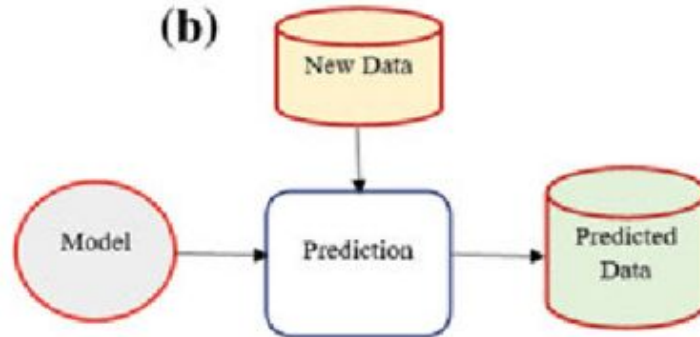


# FASE DE APRENDIZADO E FASE DE TESTE

(a)



(b)



# 2. APRESENTADO BASE DE DADOS, ENTRADAS, SAÍDAS E CONCEITOS

INTRODUÇÃO A PRÁTICA

---

# ENTRADAS, SAÍDAS E CONCEITOS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	AREA_POND	Cod_UF	Nome_da_UF	Cod_meso	Nome_da_mes	Cod_micro	Nome_da_micr	Cod_RM	Nome_da_RM	Cod_municipio	Nome_do_mun	Soma_V002.x	Soma_V023.x	Soma_V037	Soma_V038	Soma_V039	Soma_V040	Soma
2	350010500300	35	São Paulo	3508	Presidente Pru	35035	Adamantina	0	Municípios não	3500105	ADAMANTINA	4010	1	1	2	0	0	
3	350010500300	35	São Paulo	3508	Presidente Pru	35035	Adamantina	0	Municípios não	3500105	ADAMANTINA	6608	8	2	2	0	1	
4	350020400100	35	São Paulo	3501	São José do Rio	35004	São José do Rio	0	Municípios não	3500204	ADOLFO	1081	0	0	3	0	0	
5	350030300300	35	São Paulo	3507	Campinas	35029	Pirassununga	0	Municípios não	3500303	AGUAÍ	8953	9	80	12	0	12	
6	350040200100	35	São Paulo	3507	Campinas	35030	São João da Boa	0	Municípios não	3500402	ÁGUAS DA PRAT	2257	1	20	4	4	0	
7	350050100100	35	São Paulo	3507	Campinas	35033	Amparo	0	Municípios não	3500501	ÁGUAS DE LINDI	5542	5	219	40	8	0	
8	350055000100	35	São Paulo	3504	Bauru	35022	Avaré	0	Municípios não	3500550	ÁGUAS DE SANT	1432	0	11	0	0	0	
9	350060000100	35	São Paulo	3506	Piracicaba	35028	Piracicaba	0	Municípios não	3500600	ÁGUAS DE SÃO F	990	0	0	0	0	0	
10	350070900300	35	São Paulo	3504	Bauru	35020	Bauru	0	Municípios não	3500709	AGUDOS	4015	11	5	2	2	1	
11	350070900300	35	São Paulo	3504	Bauru	35020	Bauru	0	Municípios não	3500709	AGUDOS	5757	1	0	5	2	2	
12	350075800100	35	São Paulo	3511	Itapetininga	35042	Itapetininga	0	Municípios não	3500758	ALAMBARI	1134	0	12	1	0	0	
13	350080800100	35	São Paulo	3508	Presidente Pru	35036	Presidente Pru	0	Municípios não	3500808	ALFREDO MARC	1132	0	1	10	0	0	
14	350090700100	35	São Paulo	3501	São José do Rio	35004	São José do Rio	0	Municípios não	3500907	ALTAIR	944	0	0	0	0	0	
15	350100400100	35	São Paulo	3502	Ribeirão Preto	35015	Batatais	0	Municípios não	3501004	ALTINÓPOLIS	4298	2	2	0	0	0	
16	350110300100	35	São Paulo	3503	Araçatuba	35018	Birigui	0	Municípios não	3501103	ALTO ALEGRE	1096	0	2	4	0	0	
17	350115200100	35	São Paulo	3512	Macro Metropc	35046	Sorocaba	0	Municípios não	3501152	ALUMÍNIO	4176	0	3	0	0	0	
18	350120200100	35	São Paulo	3501	São José do Rio	35003	Votuporanga	0	Municípios não	3501202	ÁLVARES FLOREI	977	0	0	1	0	0	
19	350130100100	35	São Paulo	3508	Presidente Pru	35036	Presidente Pru	0	Municípios não	3501301	ÁLVARES MACH	6638	4	255	55	4	3	
20	350140000100	35	São Paulo	3509	Marília	35038	Marília	0	Municípios não	3501400	ÁLVARO DE CAR	909	1	0	1	0	0	
21	350150900100	35	São Paulo	3509	Marília	35038	Marília	0	Municípios não	3501509	ALVINLÂNDIA	822	1	1	0	0	0	
22	350160800500	35	São Paulo	3507	Campinas	35032	Campinas	22	RM Campinas	3501608	AMERICANA	4108	0	98	0	0	0	
23	350160800500	35	São Paulo	3507	Campinas	35032	Campinas	22	RM Campinas	3501608	AMERICANA	7755	2	2	0	1	0	
24	350160800500	35	São Paulo	3507	Campinas	35032	Campinas	22	RM Campinas	3501608	AMERICANA	3838	3	56	7	1	0	
25	350160800500	35	São Paulo	3507	Campinas	35032	Campinas	22	RM Campinas	3501608	AMERICANA	10143	3	65	3	0	2	
26	350160800500	35	São Paulo	3507	Campinas	35032	Campinas	22	RM Campinas	3501608	AMERICANA	7674	7	2	2	0	0	
27	350160800500	35	São Paulo	3507	Campinas	35032	Campinas	22	RM Campinas	3501608	AMERICANA	10672	4	7	7	1	1	



# ENTRADAS, SAÍDAS E CONCEITOS

AREA_POND	Soma_V002	Soma_V023	Soma_V037	...	DÉFICIT TOTAL
3500105003001	4010	1	1	...	394
3500105003002	6608	8	2	...	705
3500204001001	1081	0	0	...	112
3500303003001	8953	9	80	...	1300
...	...	...	...	...	...

# ENTRADAS, SAÍDAS E CONCEITOS

AREA_POND	Soma_V002	Soma_V023	Soma_V037	...	DÉFICIT TOTAL
3500105003001	4010	1	1	...	394
3500105003002	6608	8	2	...	705
3500204001001	1081	0	0	...	112
3500303003001	8953	9	80	...	1300
...	...	...	...	...	...

AREA_POND	3500105003001				
Setor_Cens	Soma_V002	Soma_V023	Soma_V037	...	DÉFICIT TOTAL
1000000	20	1	0	...	30
1000001	16	0	0	...	67
1000002	40	0	0	...	52
1000003	0	0	1	...	14
...	...	...	...	...	...

(Valores meramente ilustrativos!!!)

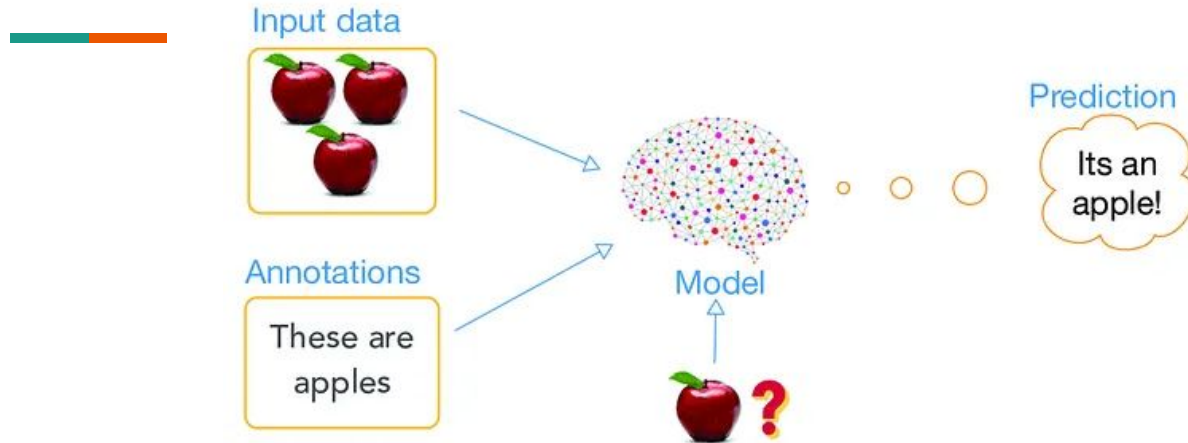
# 3. APRENDIZADO SUPERVISIONADO E NÃO SUPERVISIONADO

DESTRINCHANDO CONCEITOS

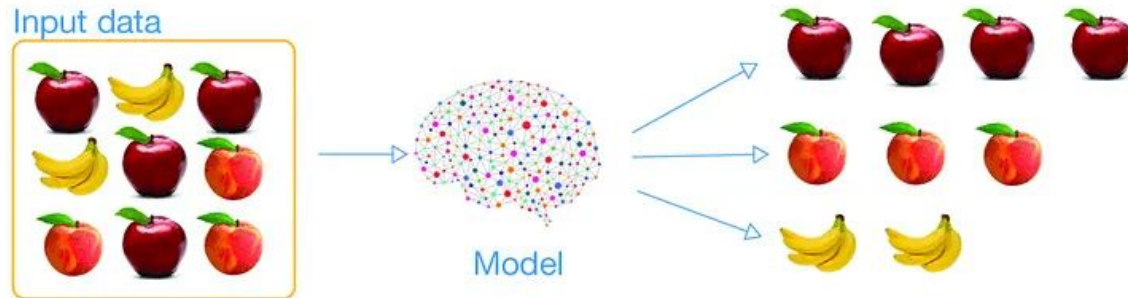


# SUPERVISIONADO X NÃO SUPERVISIONADO

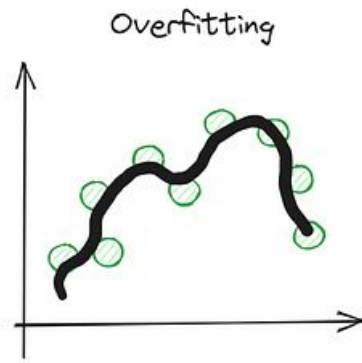
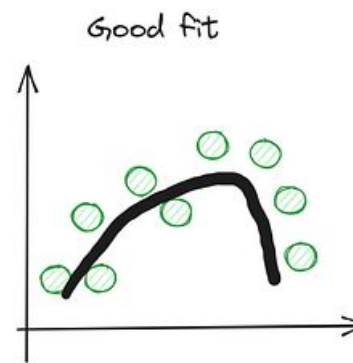
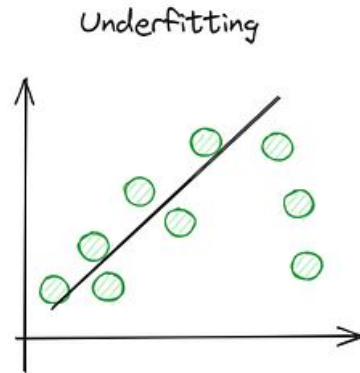
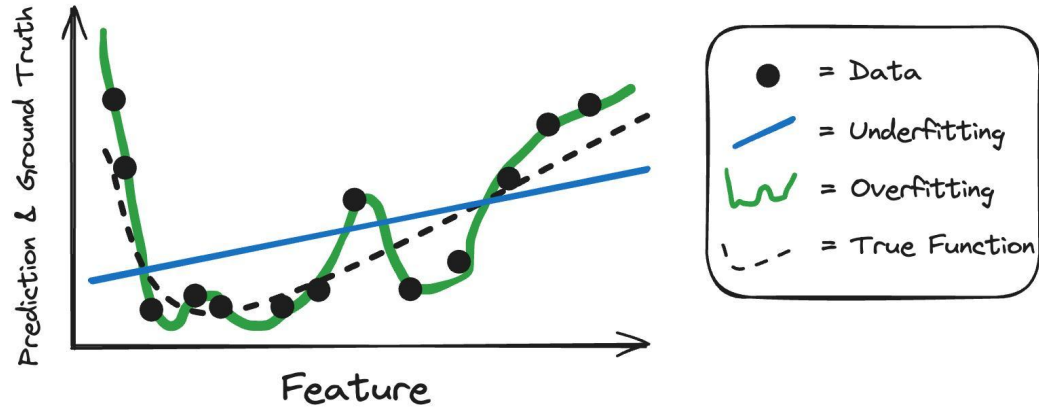
supervised learning



unsupervised learning



# SOBREAJUSTE

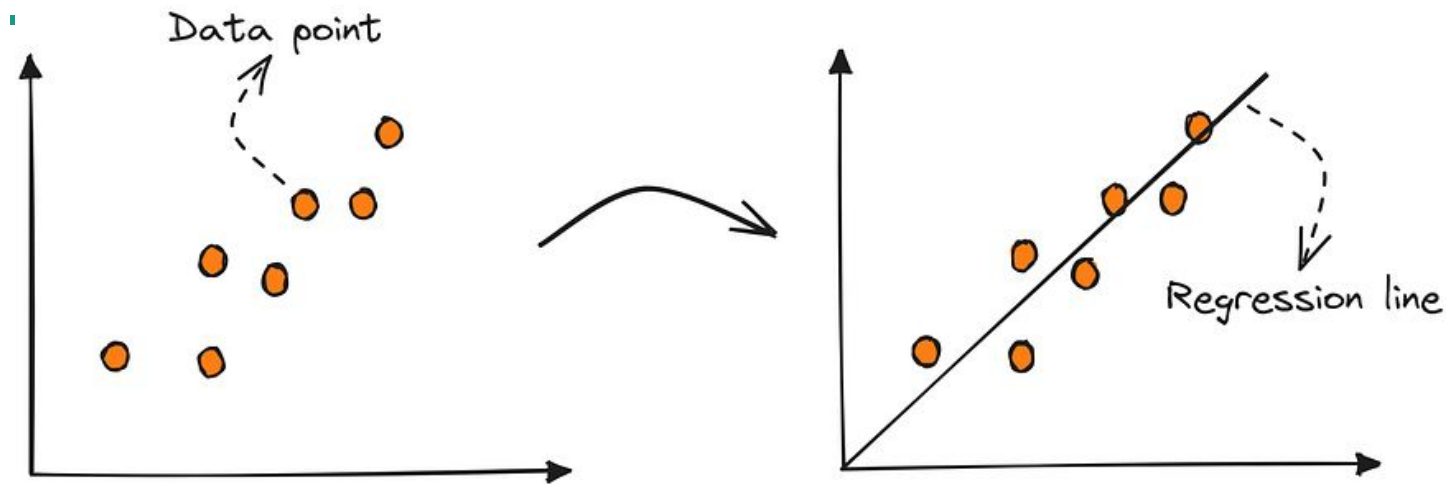


# 4. EXPLORANDO TIPOS DE APRENDIZADO SUPERVISIONADO

TRÊS PILARES: PODER PREDITIVO, PODER DE INTERPRETABILIDADE E CUSTO COMPUTACIONAL

---

# REGRESSÃO LINEAR

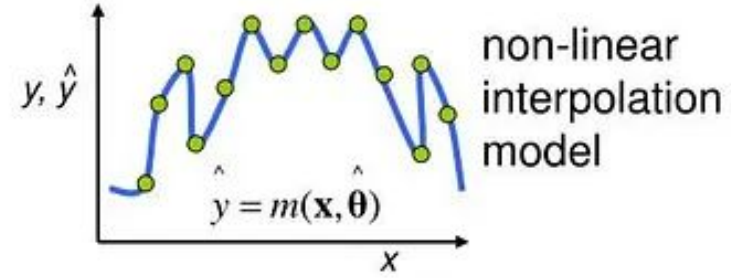
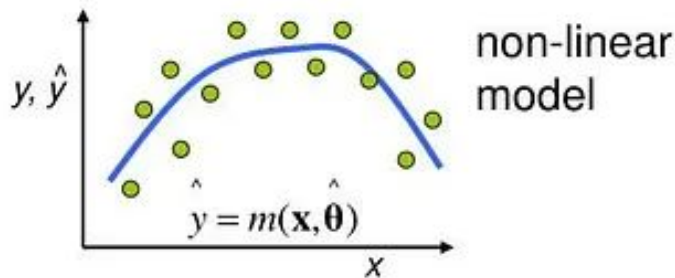
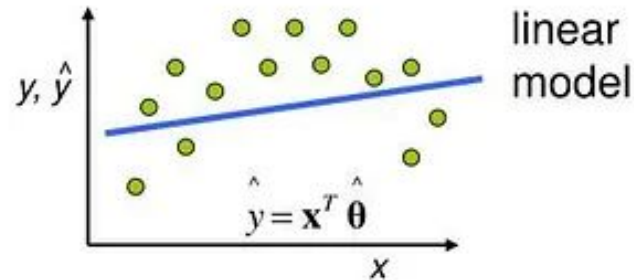
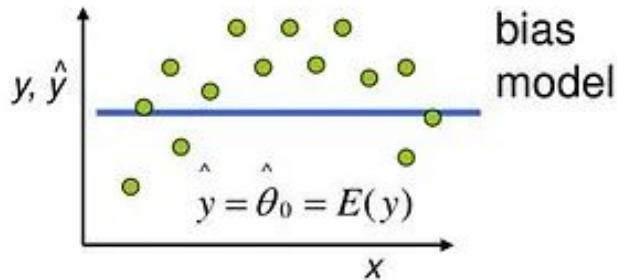


Exemplo:

Previsão =  $(\text{peso}_1 \times \text{renda}) + (\text{peso}_2 \times \text{adensamento}) + \dots$

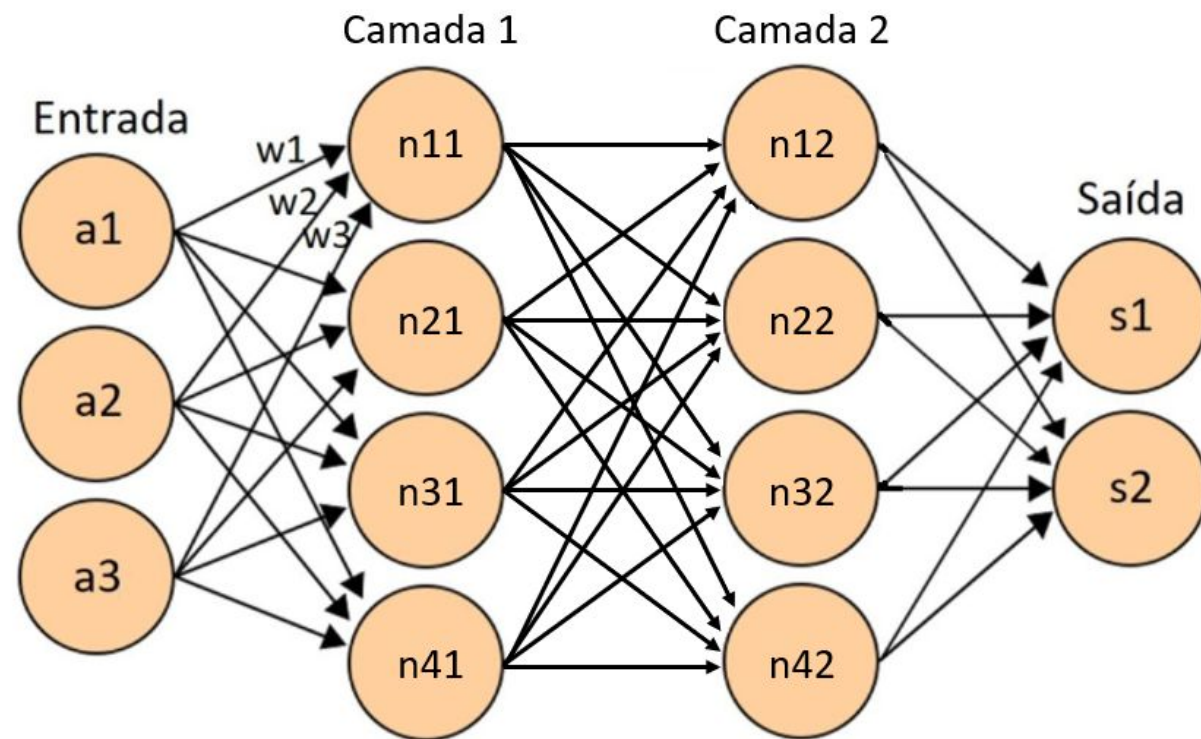
# PROBLEMAS DA REGRESSÃO LINEAR

## Non-linear Regression



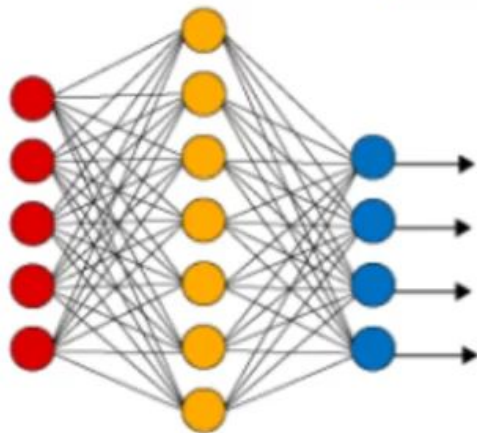


# REDES NEURAIS

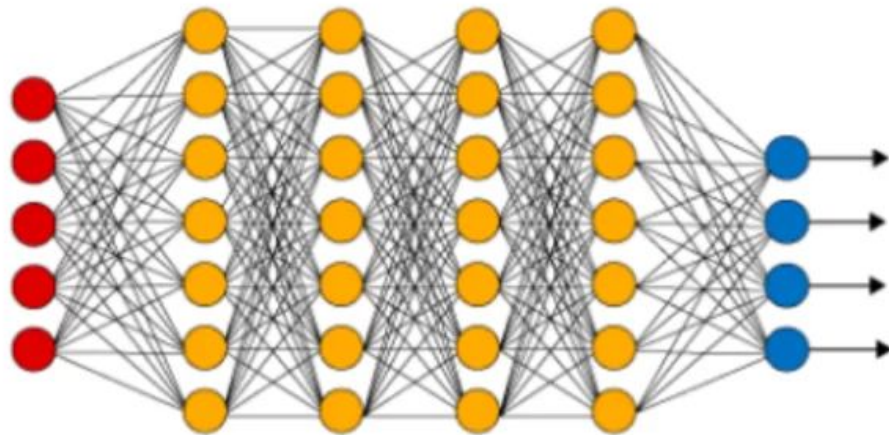


# PROBLEMAS DA REDE NEURAL

**Rede Neural Simples**

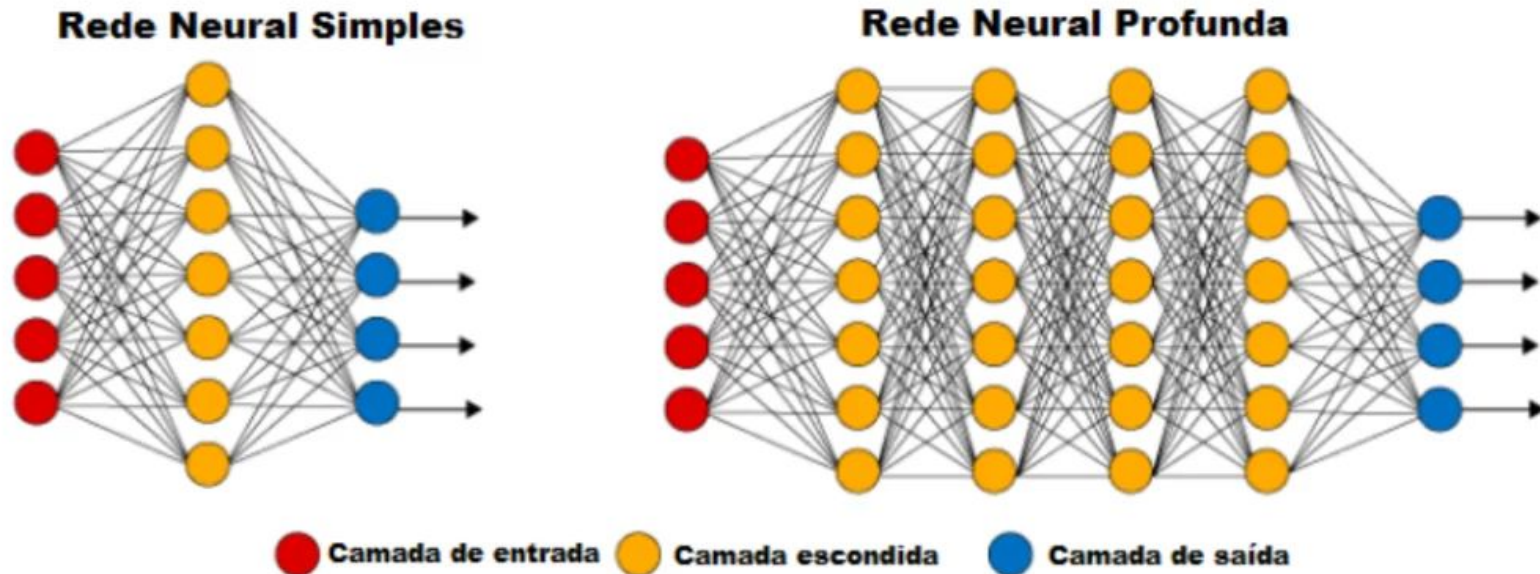


**Rede Neural Profunda**



● Camada de entrada    ● Camada escondida    ● Camada de saída

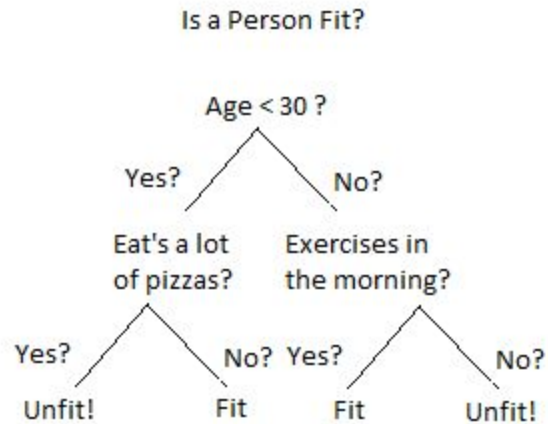
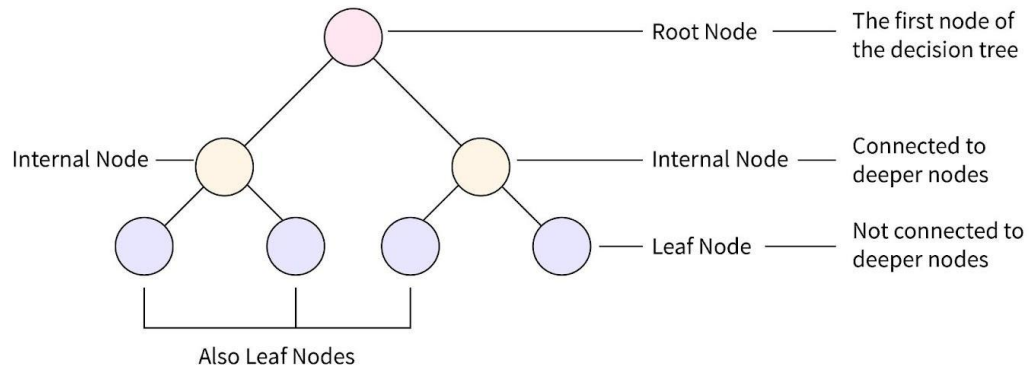
# PROBLEMAS DA REDE NEURAL



Questionamento:

A complexidade adicional, o tempo de treinamento e a perda de interpretabilidade de uma Rede Neural são justificados por uma melhoria significativa na precisão?

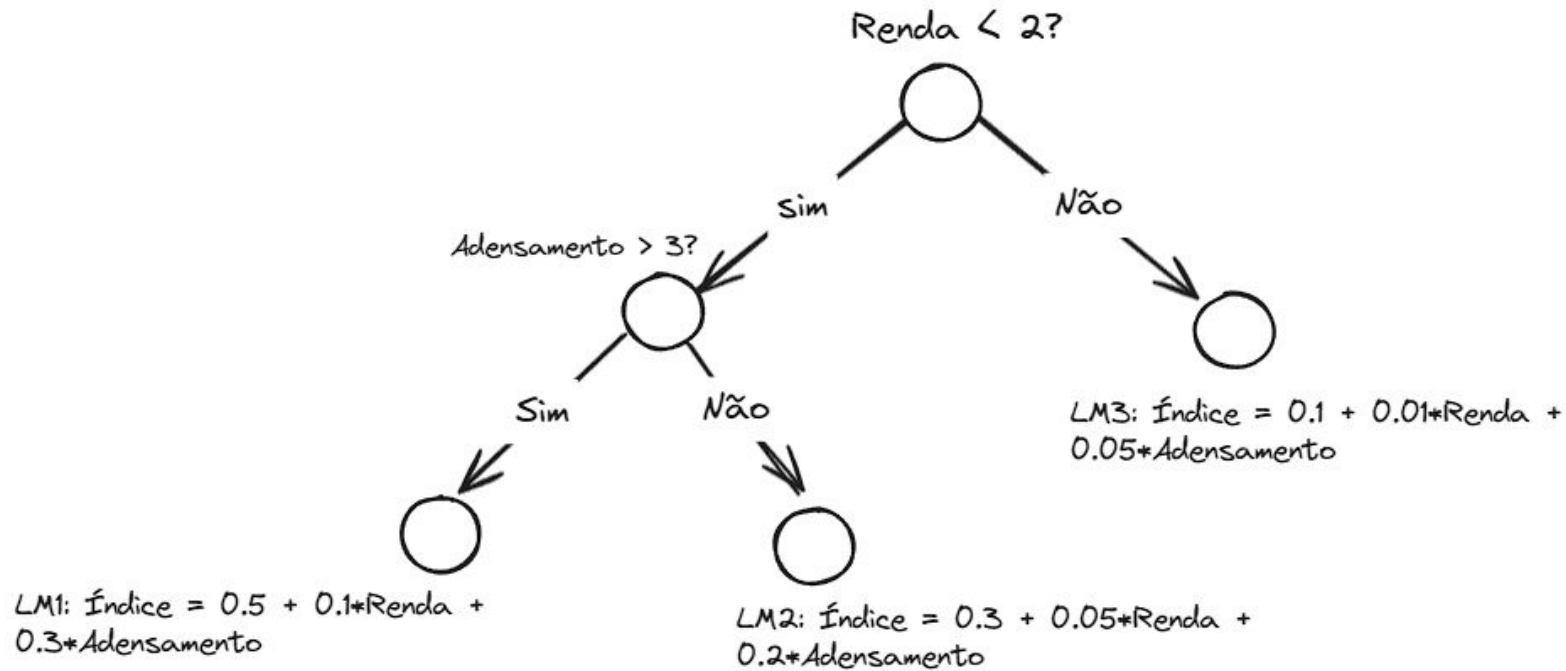
# ÁRVORE DE DECISÃO



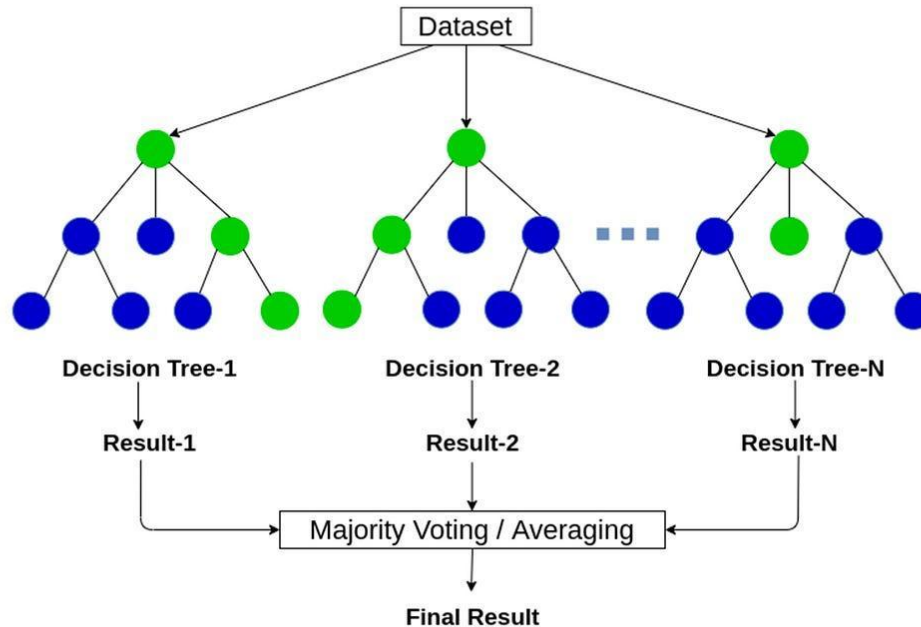
## 4. CONCLUSÕES E PRÓXIMOS PASSOS

---

# ÁRVORE DE MODELO



## Random Forest



## Comparação Direta Entre Modelos de ML

Critério	Rede Neural	Árvore de Modelo	Floresta Aleatória
<b>Precisão</b>	✓✓✓ (excelente)	✓✓	✓✓✓✓ (possivelmente a melhor)
<b>Interpretabilidade</b>	✗ (baixa, caixa preta)	✓✓✓ (vê lógica de segmentação e lógica)	✓ (sabe importância dos atributos, mas não lógica exata)
<b>Dados pequenos</b>	✗ (ruim, em poucos dados tende a sobreajuste)	✓✓✓	✓✓
<b>Valor acadêmico</b>	✓✓✓ (base teórica profunda e avanço em IA)	✓✓✓ (explica bem o déficit em vários contextos)	✓ (gera boas previsões, mas pouco contexto, difícil treinar)