

The Role of Machine Learning in Computerized Decision Making for Clinical Decision Support Systems in Healthcare

Rebecca Barter, UC Berkeley

Statistics PhD student
Data Science Fellow
at UC Berkeley

Visiting student
at UCSF





Bin Yu
UCB



Jas Sekhon
UCB



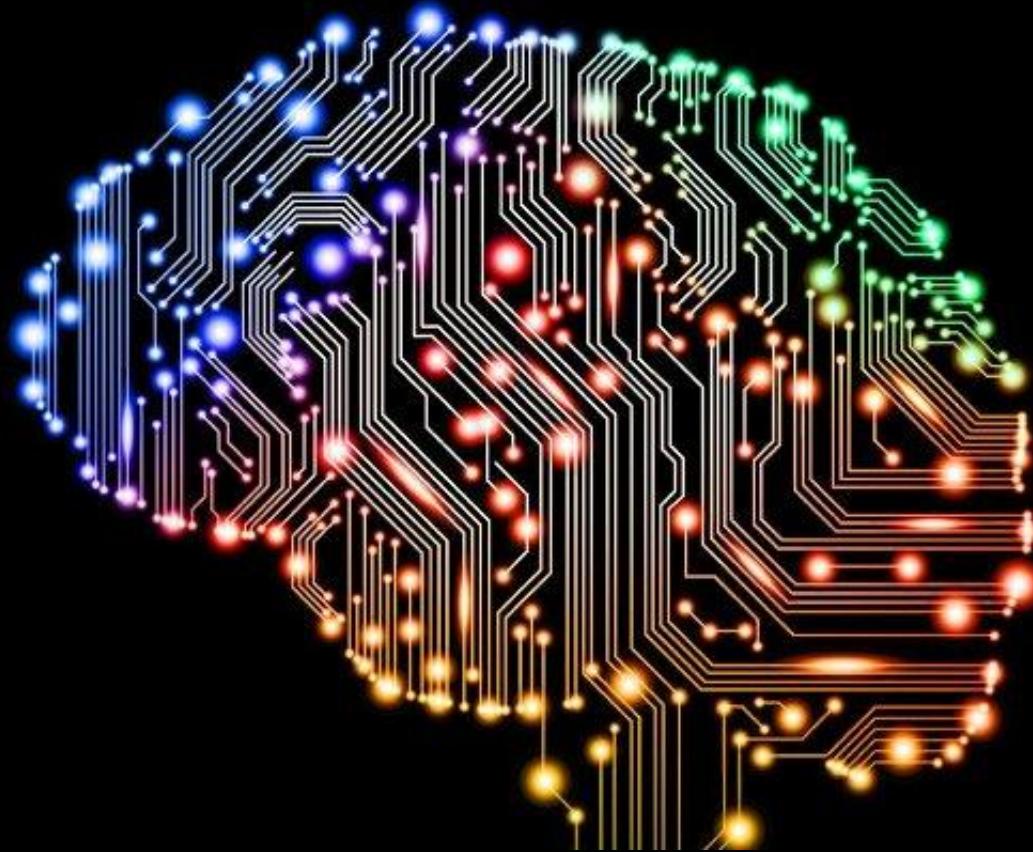
Minnie Sarwal
UCSF

My Research:

Predicting rejection in kidney transplant patients with HIV

Evaluating potential policy changes to the organ transplant waitlist

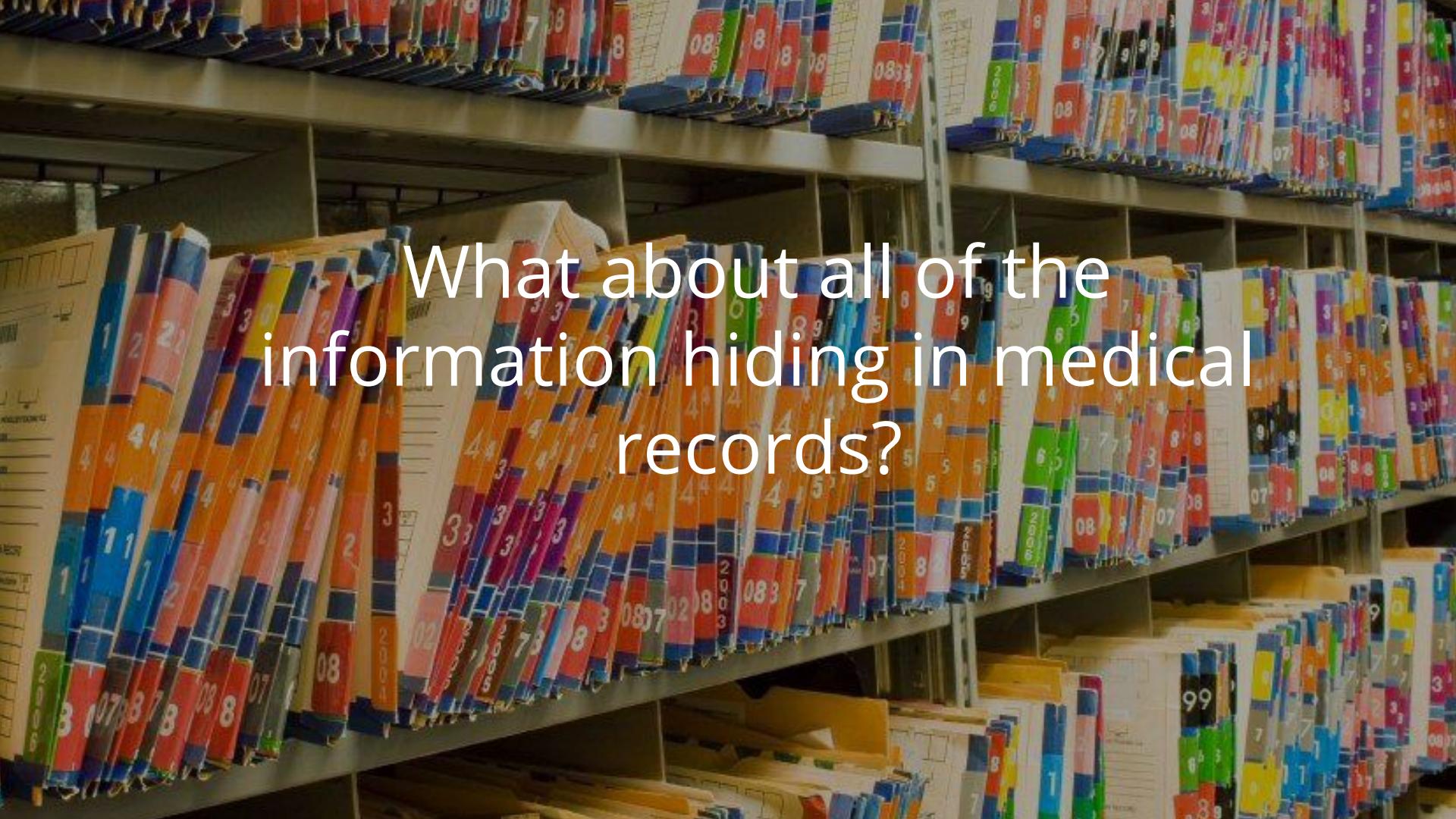
Visualizing the distribution of insulin and glucagon in the pancreas



Machine learning and Decision making

A professional photograph of a female doctor with blonde hair, wearing a white medical coat and a stethoscope. She is smiling and looking slightly to her right. In her hands, she holds a blue pen over a white document. The background shows a clinical setting with light-colored cabinets.

Doctors base decisions on:
Past experience and medical knowledge

A photograph showing a vast collection of medical records, likely from a hospital or clinic. The records are organized on multiple metal shelving units. Each record is a folder with a unique color scheme and a white tab at the top displaying a four-digit number, such as '0833'. The colors of the folders range from blues and reds to yellows, greens, and purples. The shelves are packed closely together, creating a dense wall of information.

What about all of the
information hiding in medical
records?



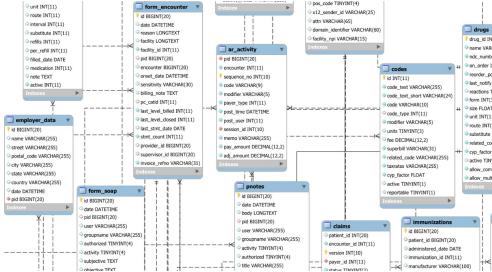
Or in the genome?

A close-up, low-angle shot of a massive stack of scientific journals or publications. The pages are tightly packed, creating a textured, layered effect. The colors of the covers are varied, with many featuring scientific illustrations, graphs, and text. The perspective is from the side, looking down the length of the stack.

In scientific publications?

A photograph of a long, narrow corridor in a data center. Both sides are filled with tall server racks, their front panels illuminated with numerous small, glowing green and yellow lights. The floor is a polished white tile. In the distance, a person stands at the far end of the corridor, appearing very small due to perspective. The overall atmosphere is one of a vast, high-tech environment.

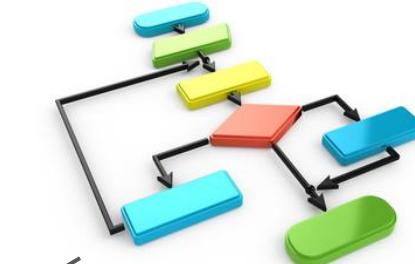
Healthcare is in the middle of a
“Big Data” Explosion



Data



Medical
knowledge



Algorithms

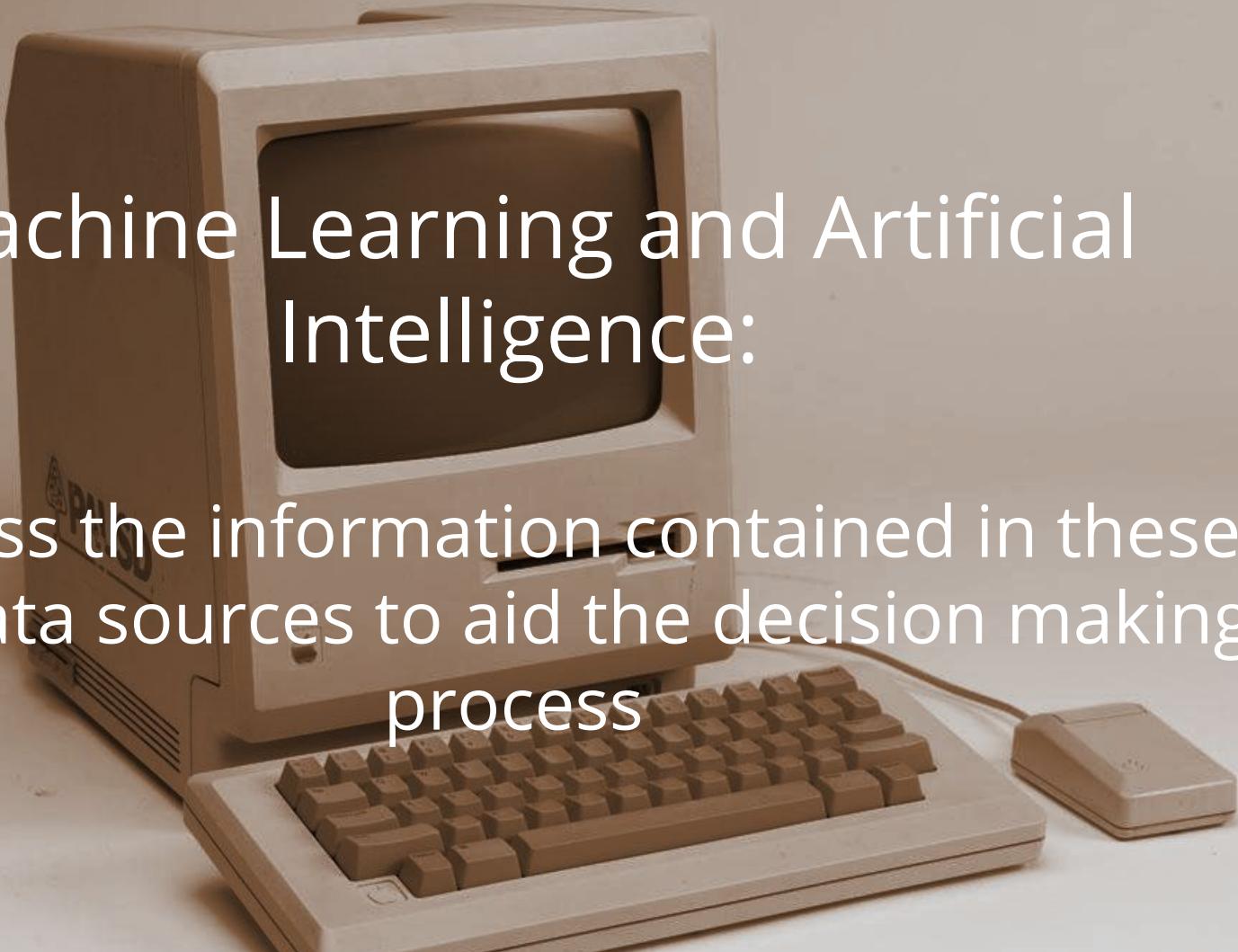
Clinical Decision Support



Automatic information, warnings and alerts

Machine Learning and Artificial Intelligence:

Harness the information contained in these vast data sources to aid the decision making process



A surgeon in green scrubs and a red stethoscope around their neck is holding a light-colored tablet computer in their right hand. In the background, other medical professionals in green scrubs are visible in a blurred hospital setting.

Predict patient diagnosis and outcomes

Predict optimal drug regimen

A close-up photograph of a variety of prescription medications, including tablets and capsules in shades of orange, yellow, green, red, and blue, spilling out of a yellow prescription bottle onto a reflective surface.

Predict costs

HEALTH INSURANCE CLAIM FORM

A. POLICYHOLDER -Insured Details

Insurance number

Date of birth

Postcode and town

Phone (+country code and local dialing code)

First name(s)/surname/title

Correspondence address

Fax (+country code and local

B. PATIENT DETAILS

Insured's or co-insured's number

First name(s)/surname/title

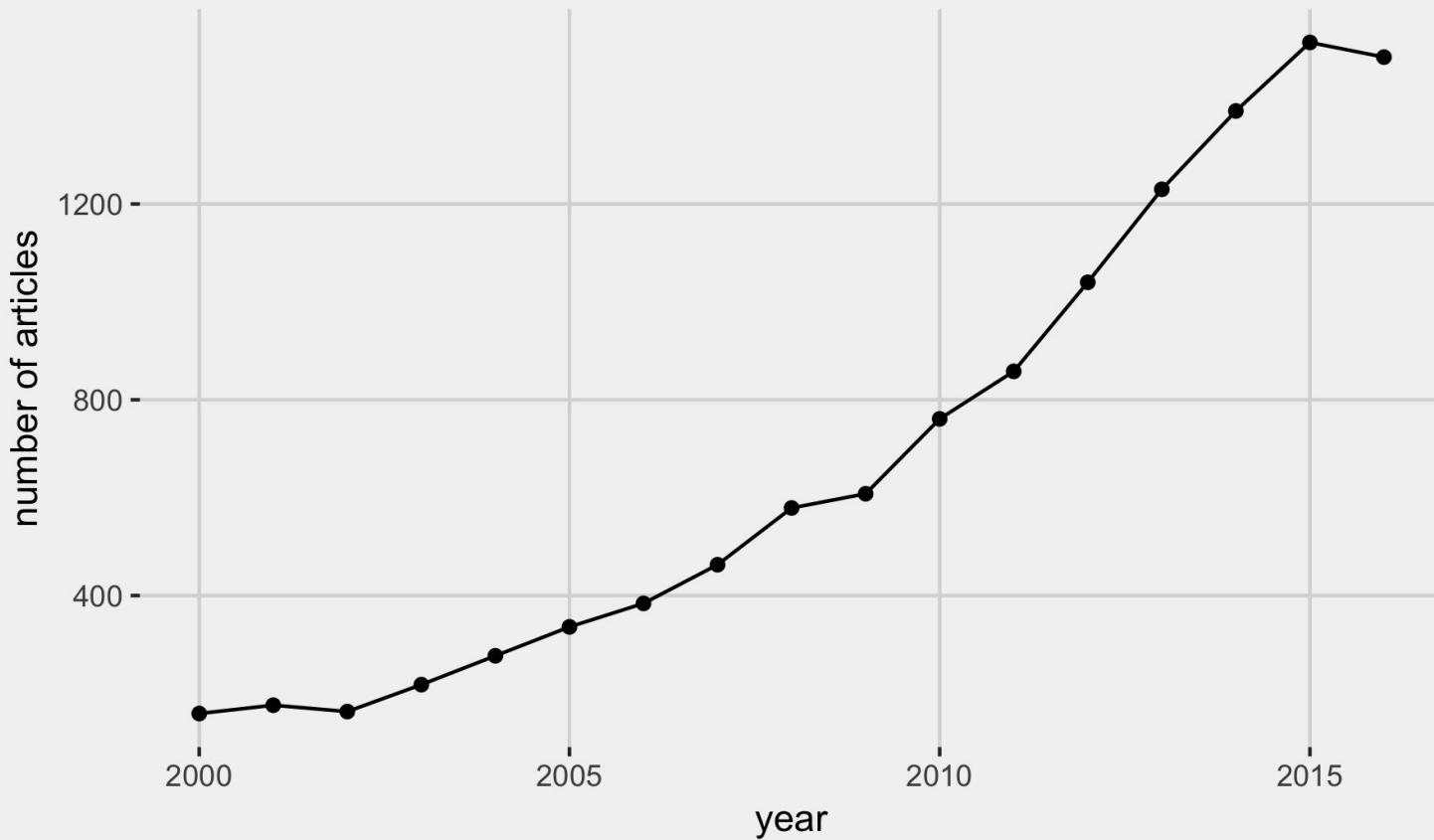
Want?

yes

claim
how it occurs



Number of articles involving 'Clinical Decision Support System' and 'Machine Learning'



Data based on Google Scholar

How does machine learning work?



How does machine learning work?

Data as input:

Text files,
spreadsheets,
databases

How does machine learning work?

Data as input:

Text files,
spreadsheets,
databases



**Abstracting the
data:**

Computer learns an
algorithmic
representation of the
data

How does machine learning work?

Data as input:

Text files,
spreadsheets,
databases

**Abstracting the
data:**

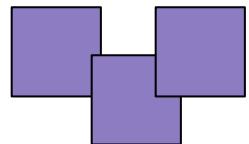
Computer learns an
algorithmic
representation of the
data

Generalization:

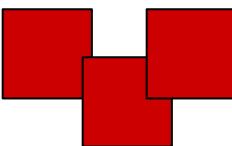
Develop an
insight. Apply
the algorithm to
new data.

Two main types of machine learning

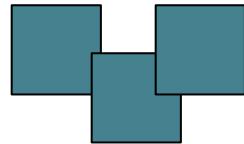
Supervised learning



Group 1



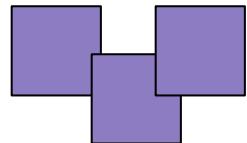
Group 2



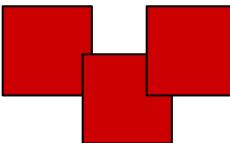
Group 3

Two main types of machine learning

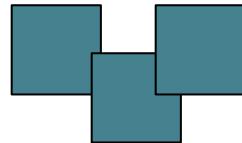
Supervised learning



Group 1



Group 2

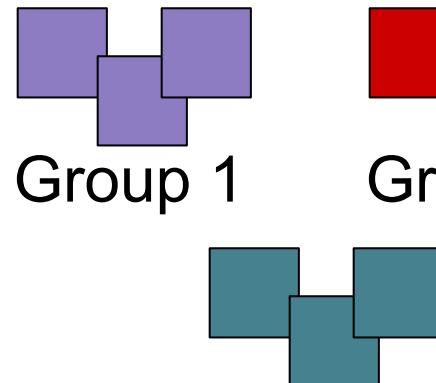


Group 3

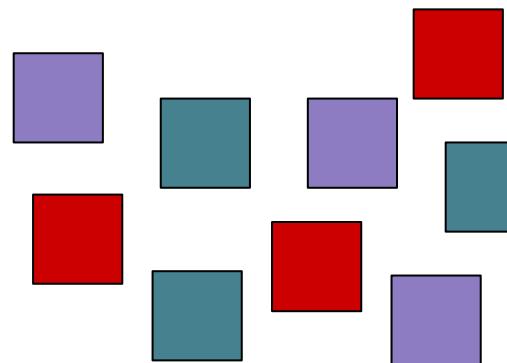
What group does fall into?

Two main types of machine learning

Supervised learning



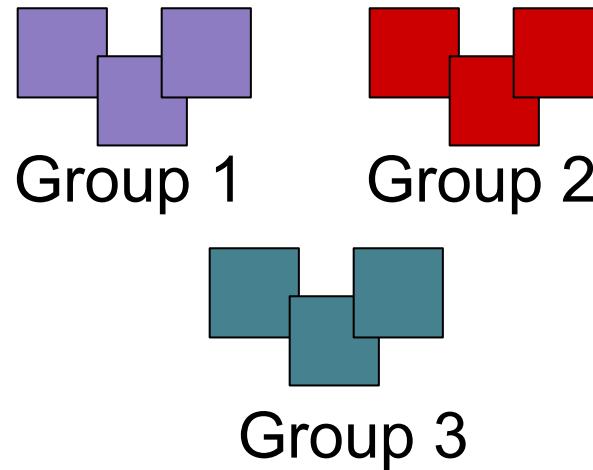
Unsupervised learning



What group does fall into?

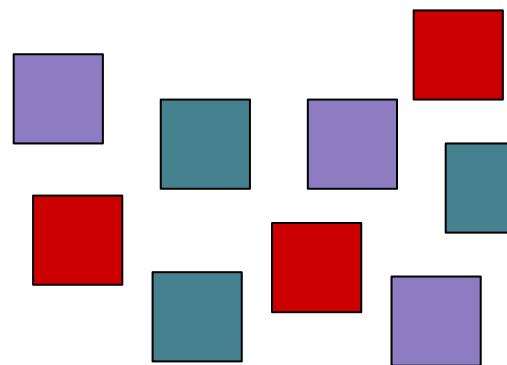
Two main types of machine learning

Supervised learning



What group does fall into?

Unsupervised learning



Are there groups in the data?

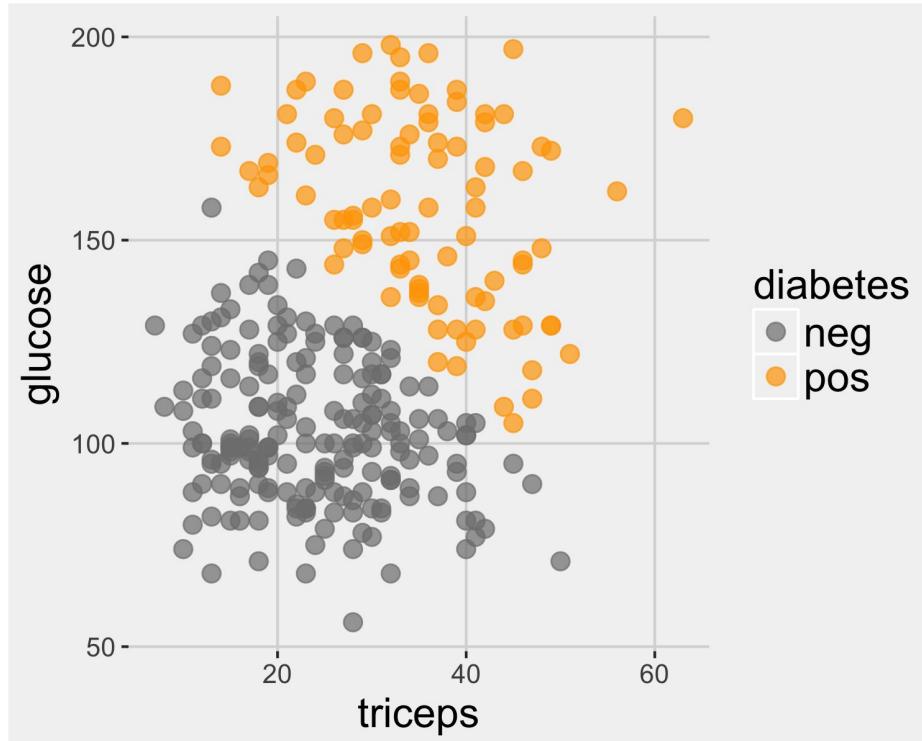
Supervised learning: predicting diabetes in pregnant Pima Indian women

Supervised learning:

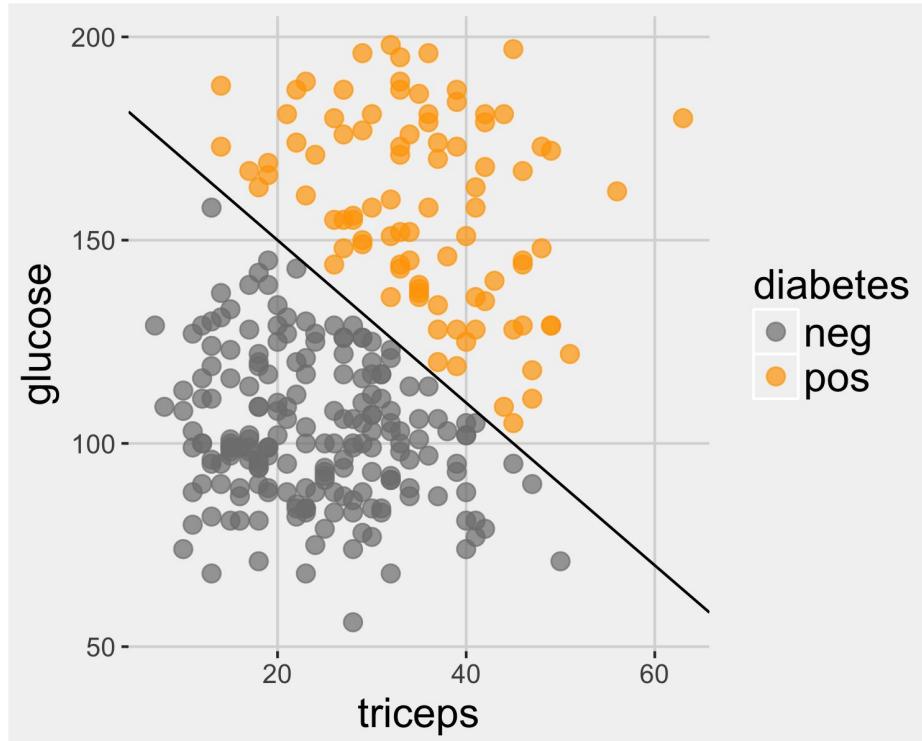
predicting diabetes in pregnant Pima Indian women

Pregnant	Glucose	Pressure	Triceps	Insulin	BMI	Age	Diabetes
1	121	78	39	74	39.0	28	?
0	181	88	44	510	43.3	26	?

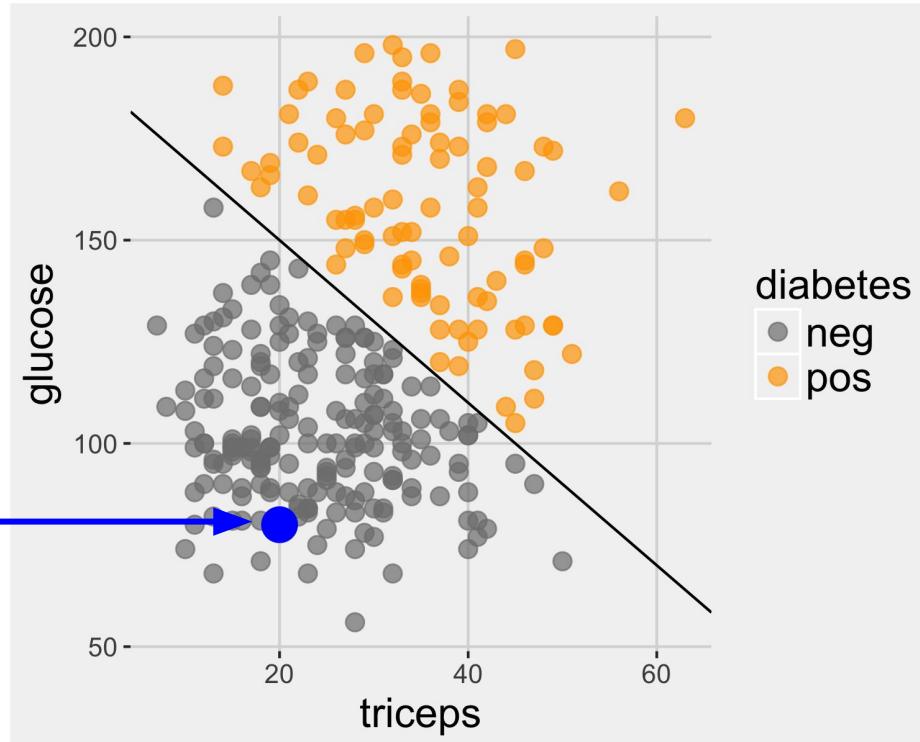
Supervised learning: predicting diabetes in pregnant Pima Indian women



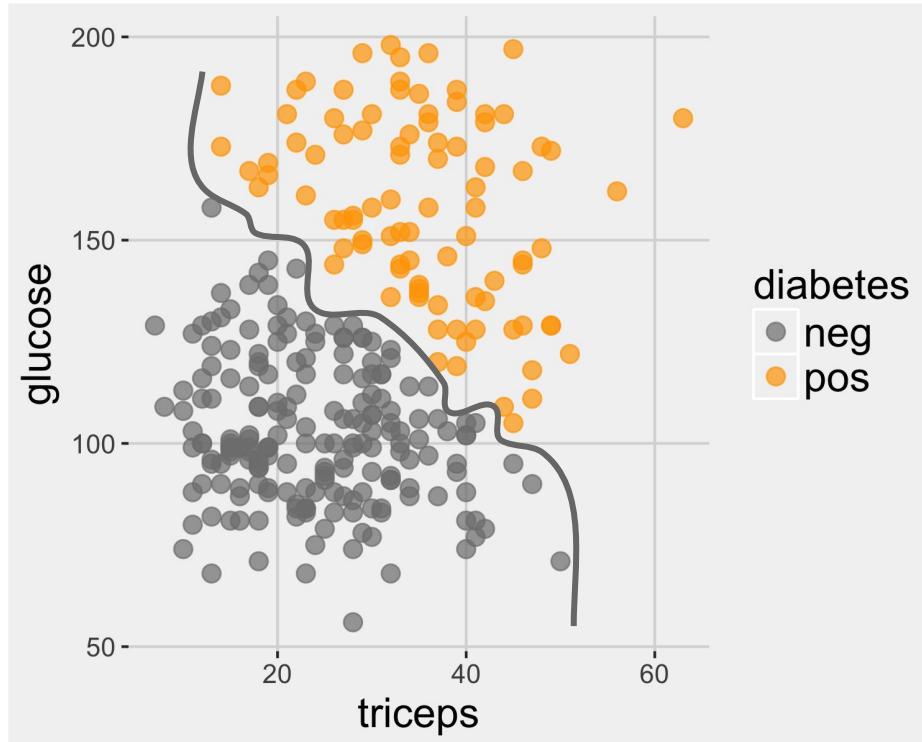
Supervised learning: predicting diabetes in pregnant Pima Indian women



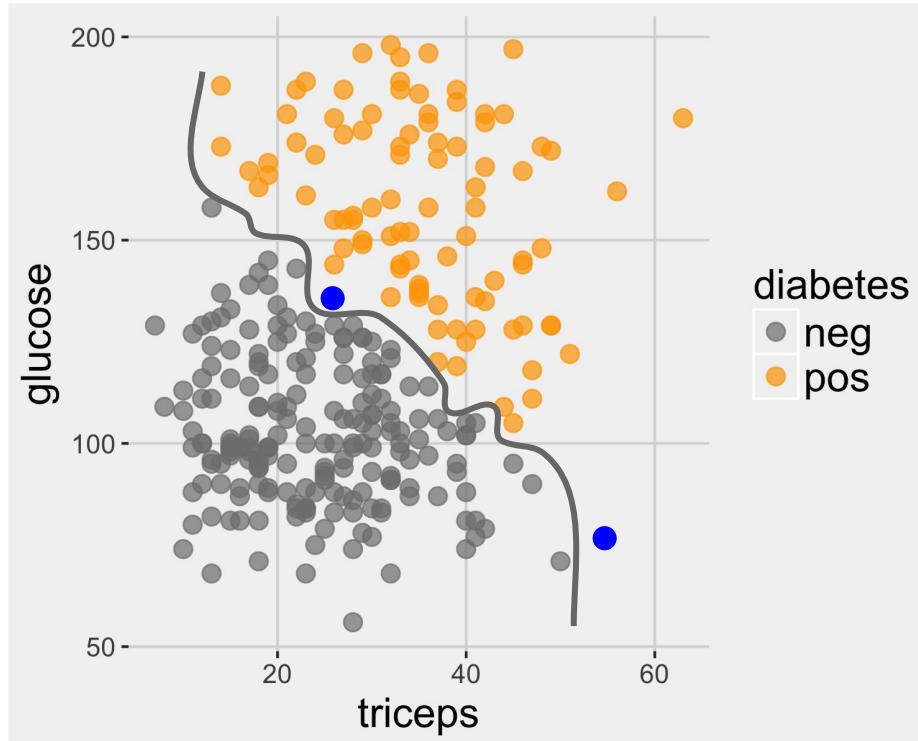
Supervised learning: predicting diabetes in pregnant Pima Indian women



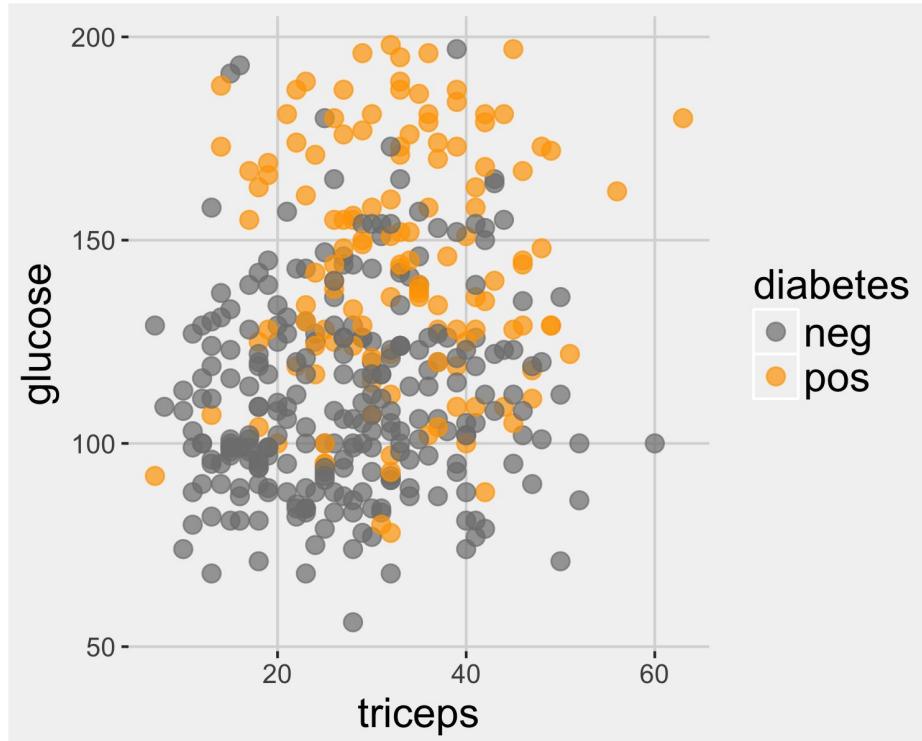
Supervised learning: predicting diabetes in pregnant Pima Indian women



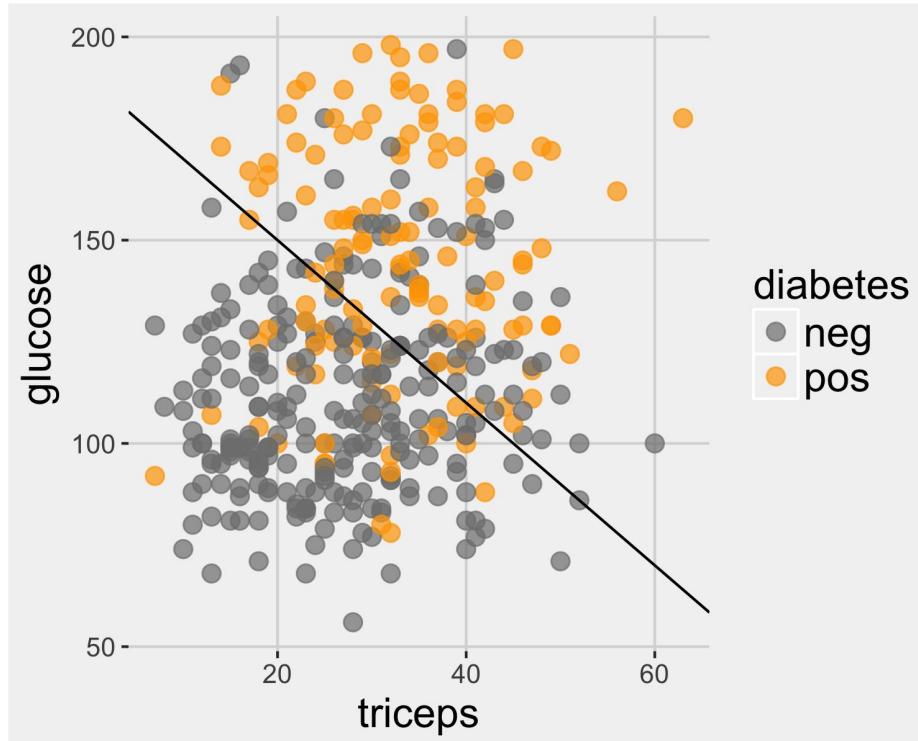
Supervised learning: predicting diabetes in pregnant Pima Indian women



Supervised learning: predicting diabetes in pregnant Pima Indian women

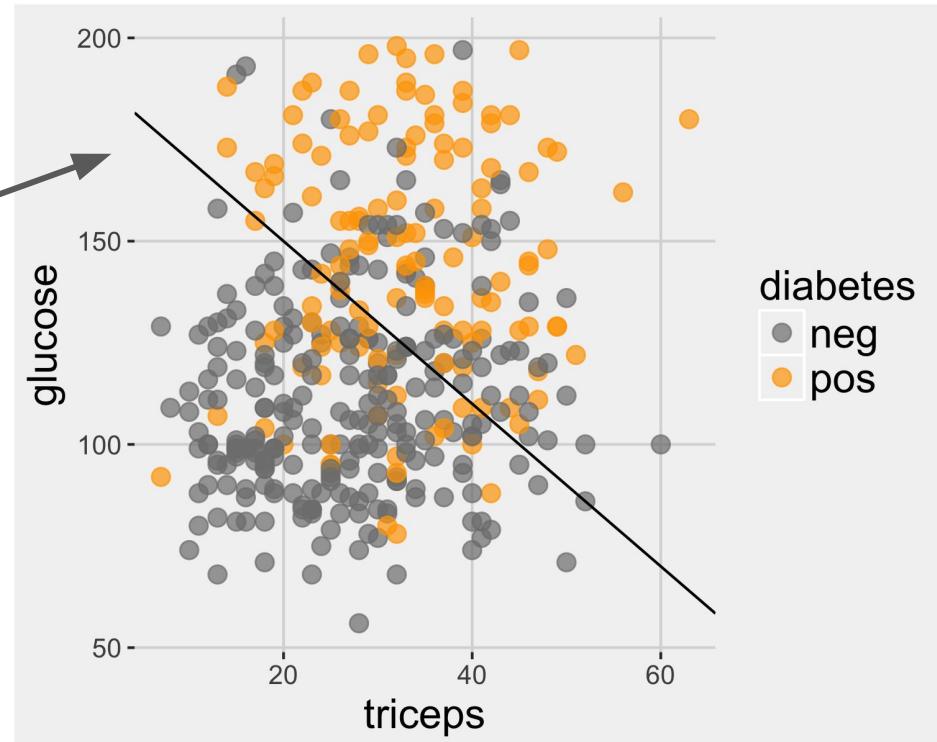


Supervised learning: predicting diabetes in pregnant Pima Indian women

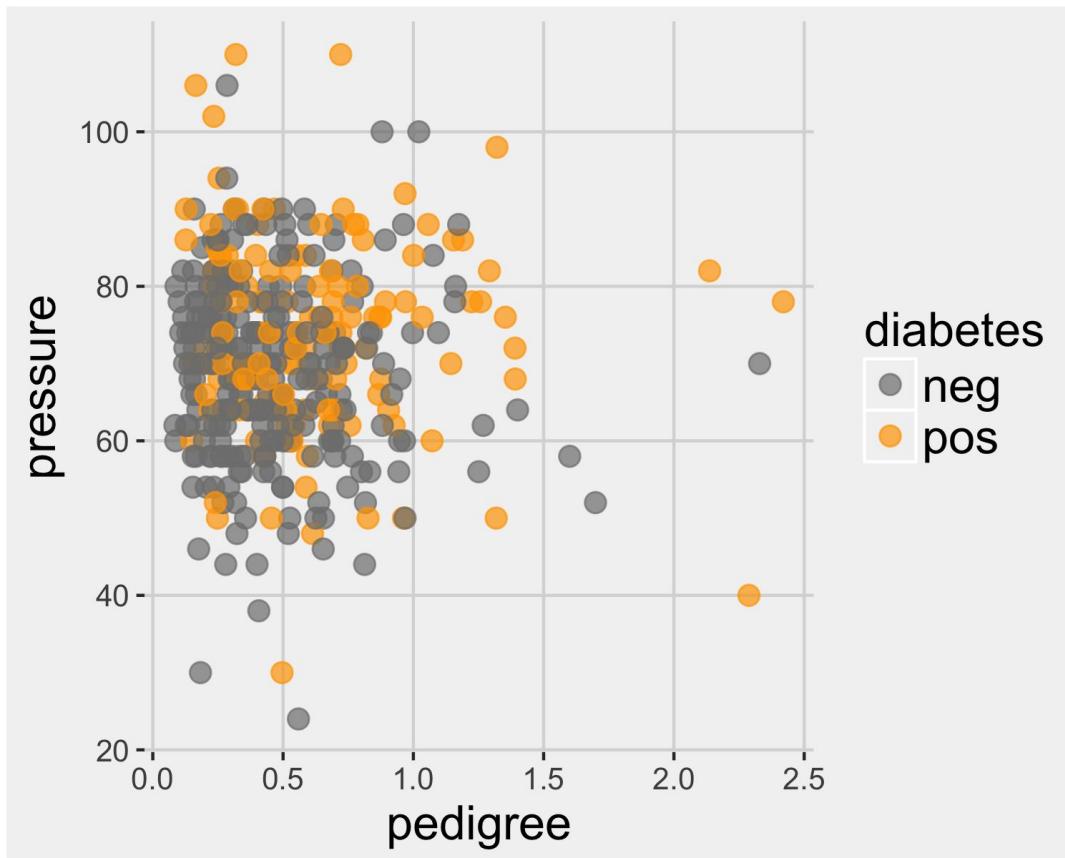


Supervised learning: predicting diabetes in pregnant Pima Indian women

Machine learning is all
about generalizing this
separating line!



The importance of meaningful features



Evaluating a machine learning model

Withhold samples

Withhold samples

Pregnant	Glucose	Pressure	Triceps	Insulin	BMI	Age	Diabetes
1	89	66	23	94	28.1	21	no
3	78	50	32	88	31.0	26	yes
2	197	70	45	543	30.5	53	yes
...

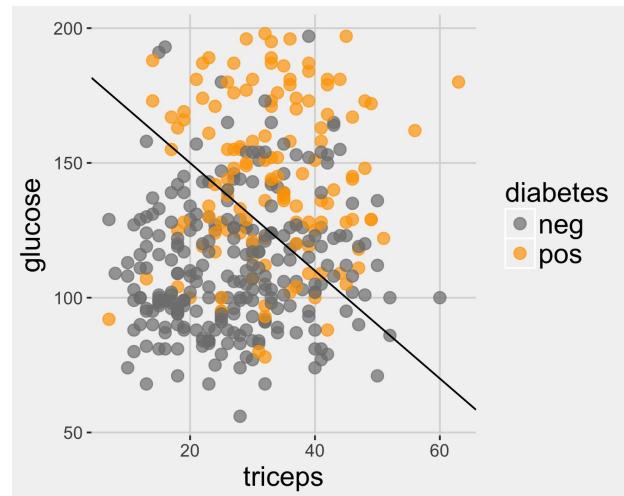
Training set

Pregnant	Glucose	Pressure	Triceps	Insulin	BMI	Age	Diabetes
0	137	40	35	168	43.1	33	yes
1	189	60	23	846	30.1	59	yes
...

Test set

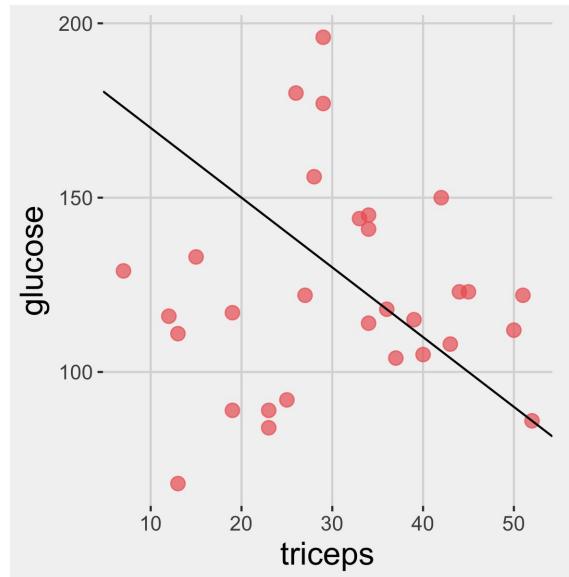
Training set: build model

Pregnant	Glucose	Pressure	Triceps	Insulin	BMI	Age	Diabetes
1	89	66	23	94	28.1	21	no
3	78	50	32	88	31.0	26	yes
2	197	70	45	543	30.5	53	yes
...



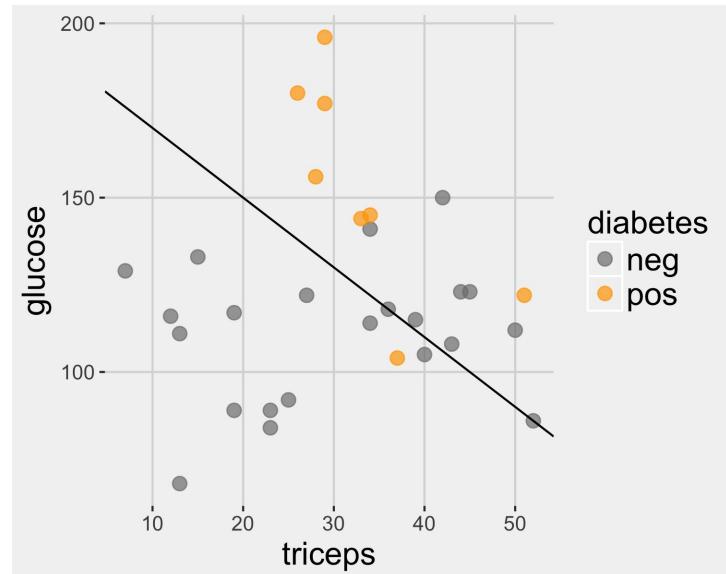
Test set: use model to predict class

Pregnant	Glucose	Pressure	Triceps	Insulin	BMI	Age	Diabetes
0	137	40	35	168	43.1	33	yes
1	189	60	23	846	30.1	59	yes
...



Test set: how well did we do?

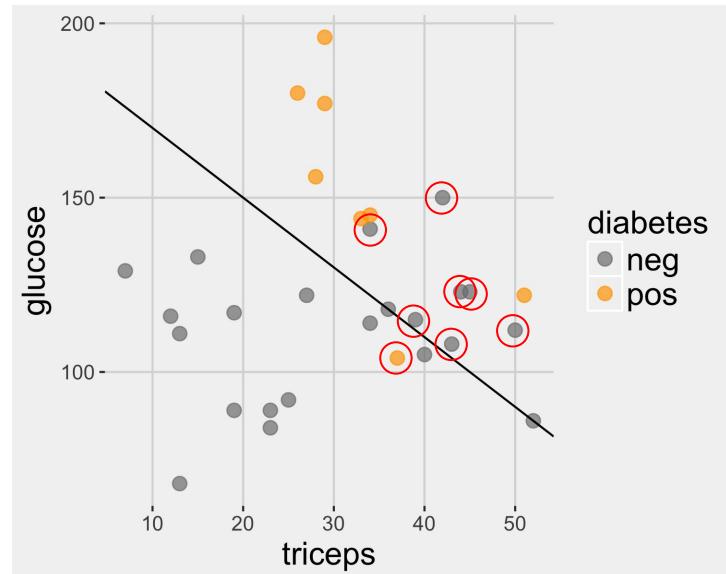
Pregnant	Glucose	Pressure	Triceps	Insulin	BMI	Age	Diabetes
0	137	40	35	168	43.1	33	yes
1	189	60	23	846	30.1	59	yes
...



Test set: how well did we do?

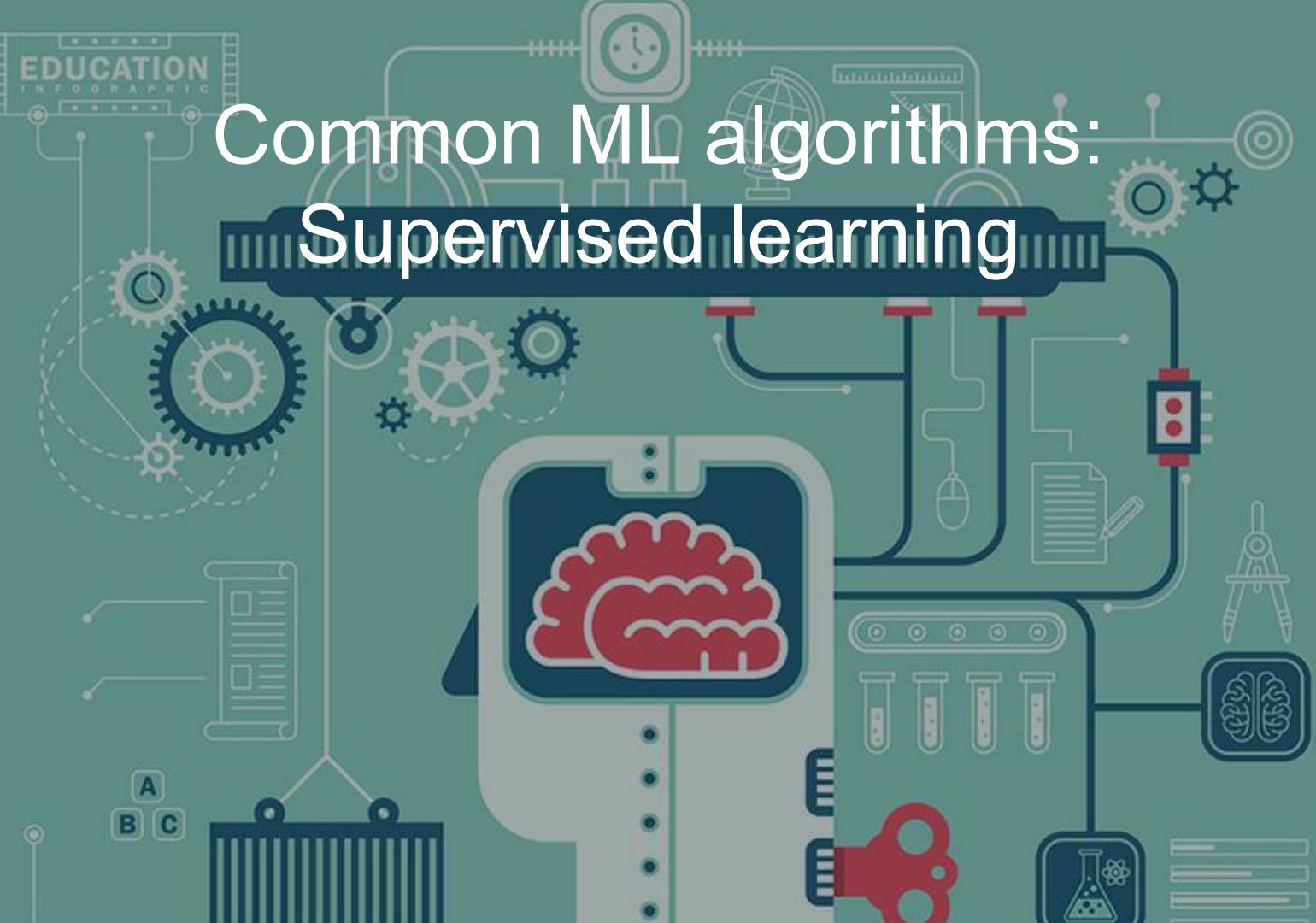
Accuracy: 73%

Pregnant	Glucose	Pressure	Triceps	Insulin	BMI	Age	Diabetes
0	137	40	35	168	43.1	33	yes
1	189	60	23	846	30.1	59	yes
...

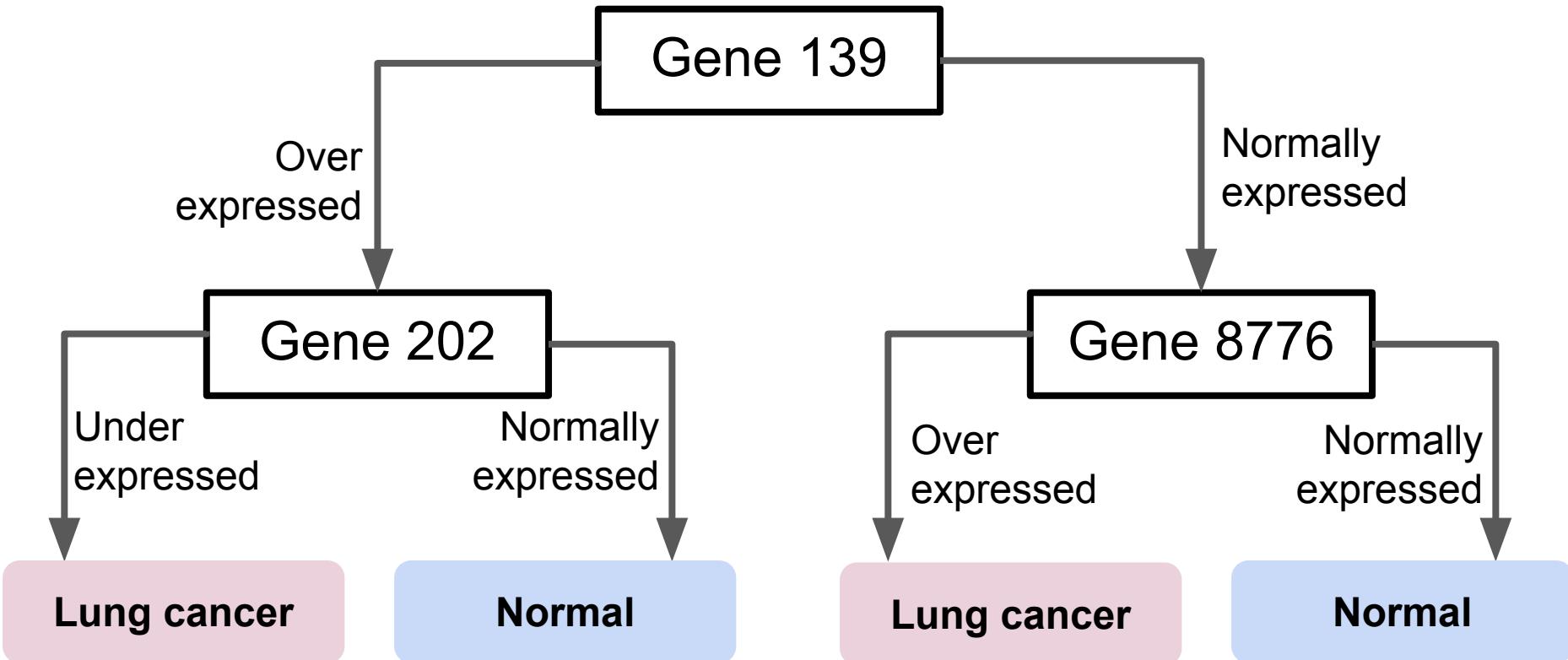




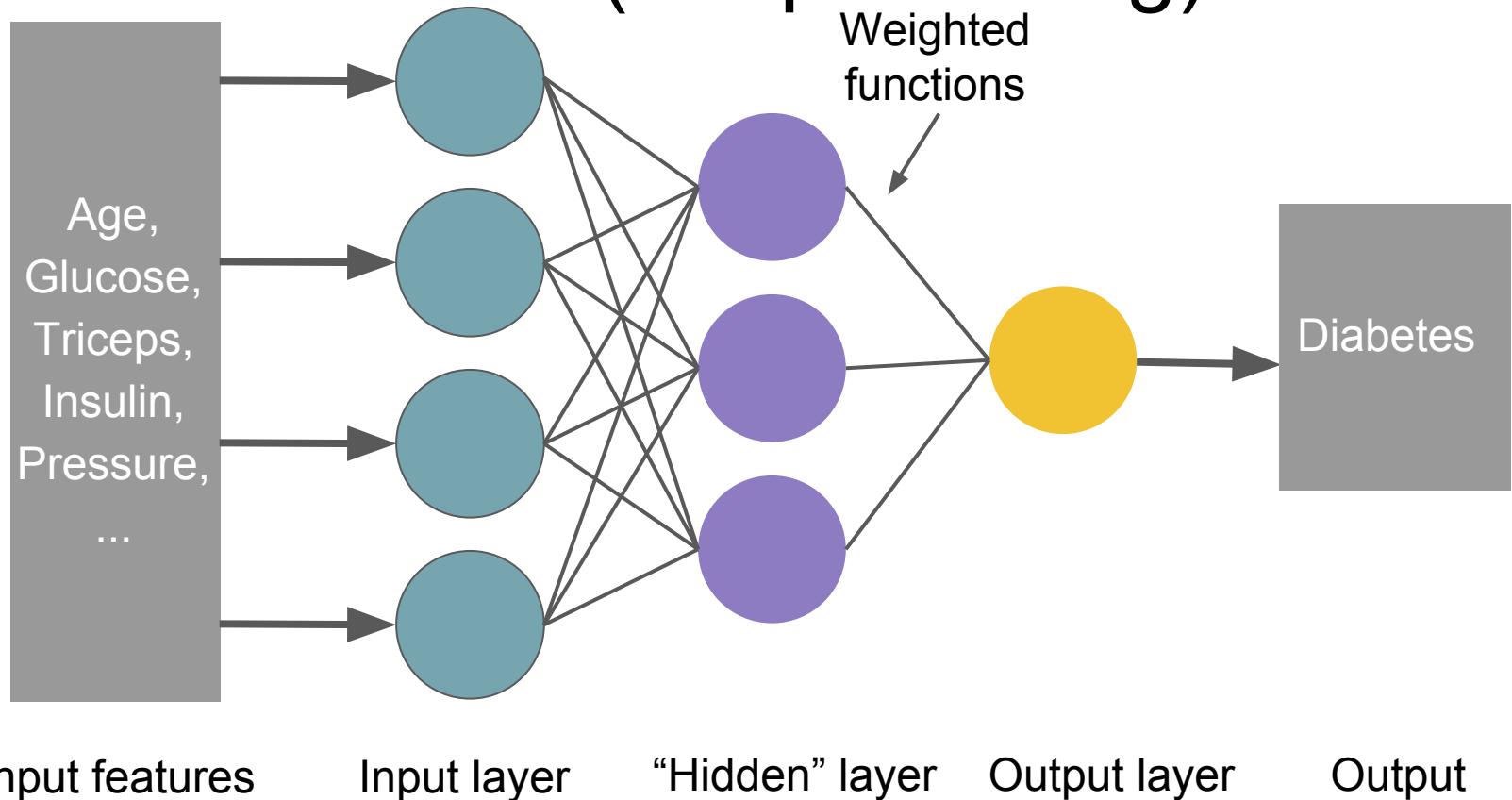
Common ML algorithms: Supervised learning



Decision trees



Neural networks (deep learning)



Neural networks (deep learning)

Age

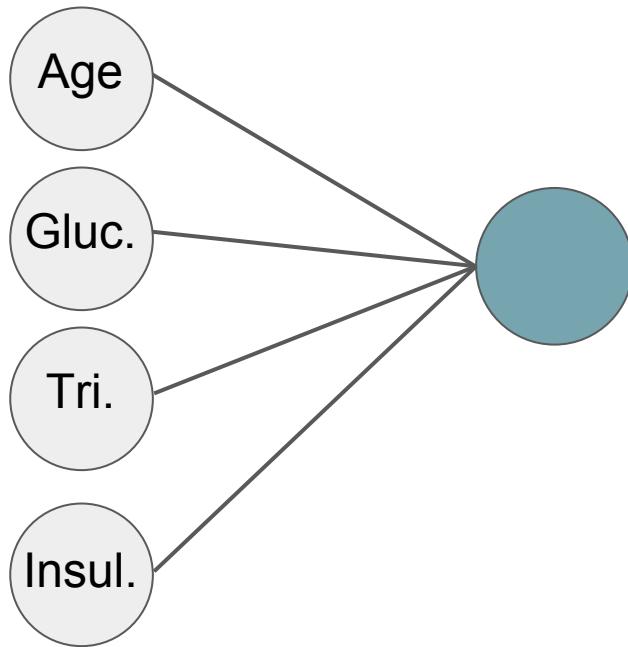
Gluc.

Tri.

Insul.

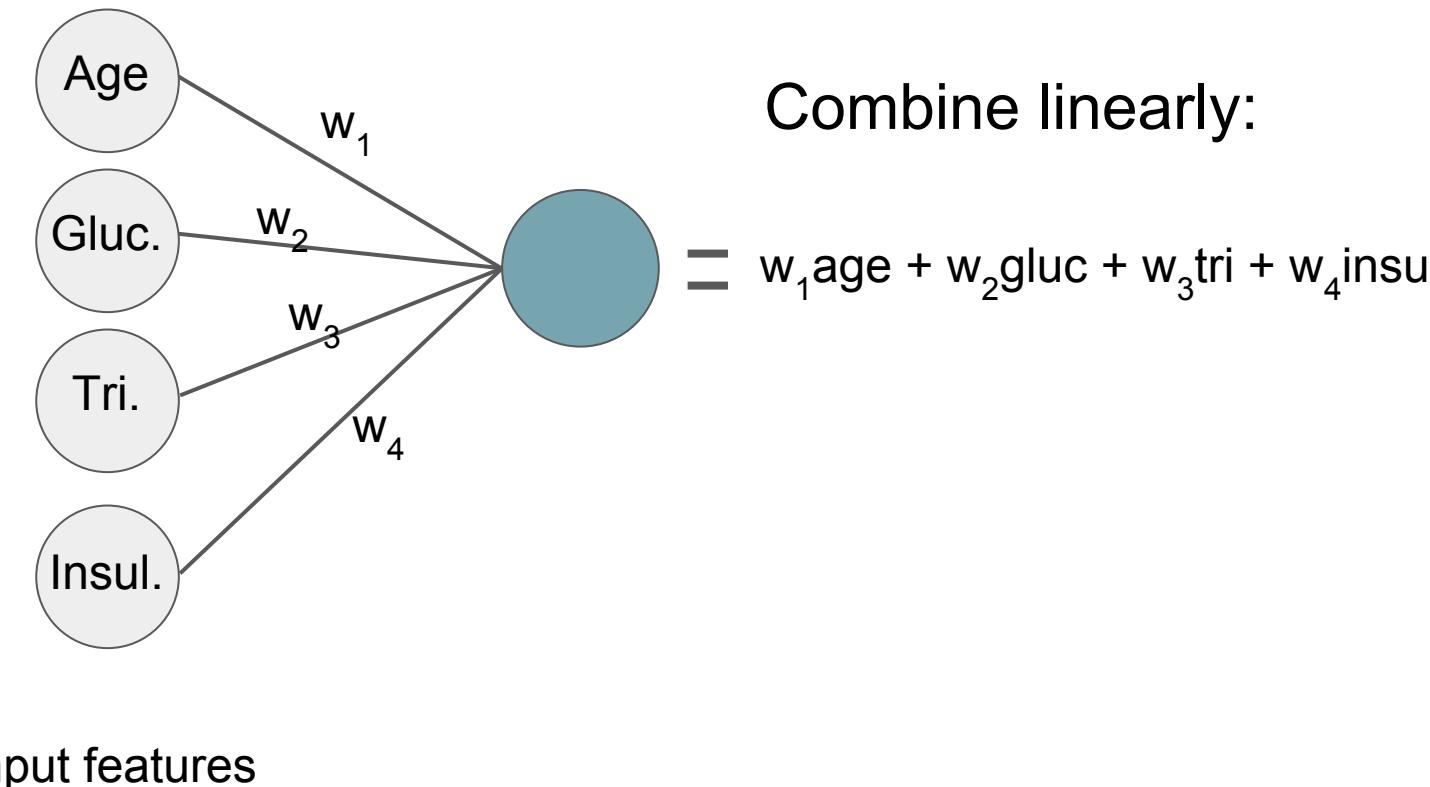
Input features

Neural networks (deep learning)

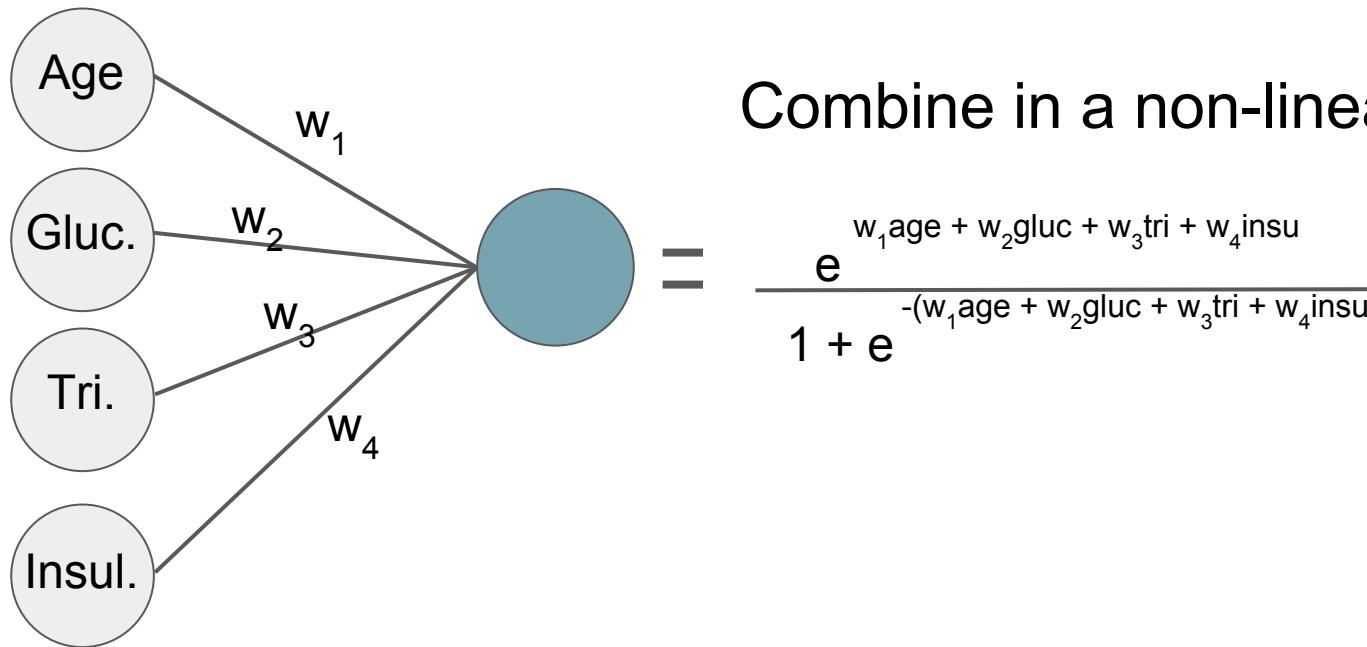


Input features

Neural networks (deep learning)

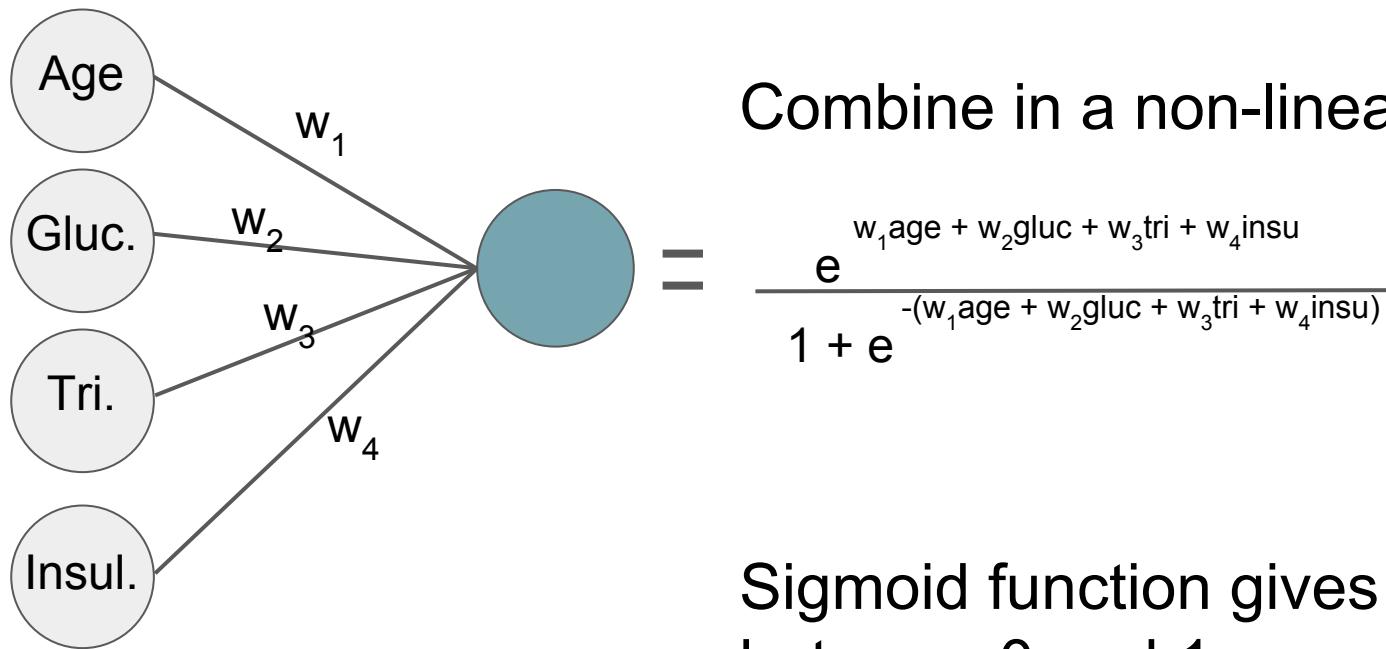


Neural networks (deep learning)



Combine in a non-linear way

Neural networks (deep learning)

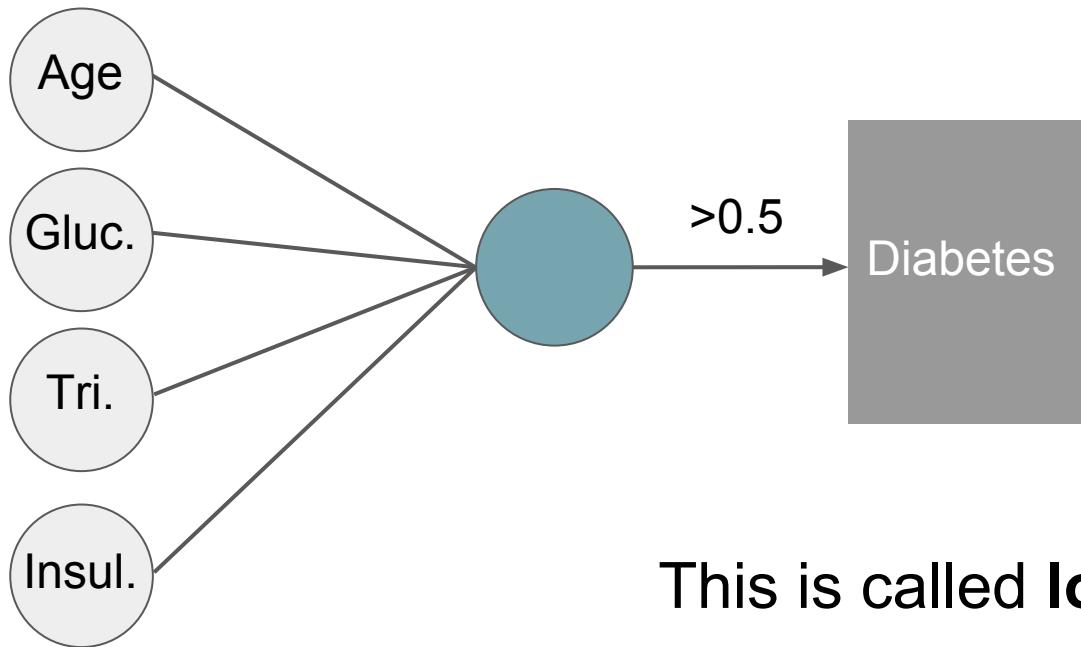


Input features

Combine in a non-linear way

Sigmoid function gives value between 0 and 1

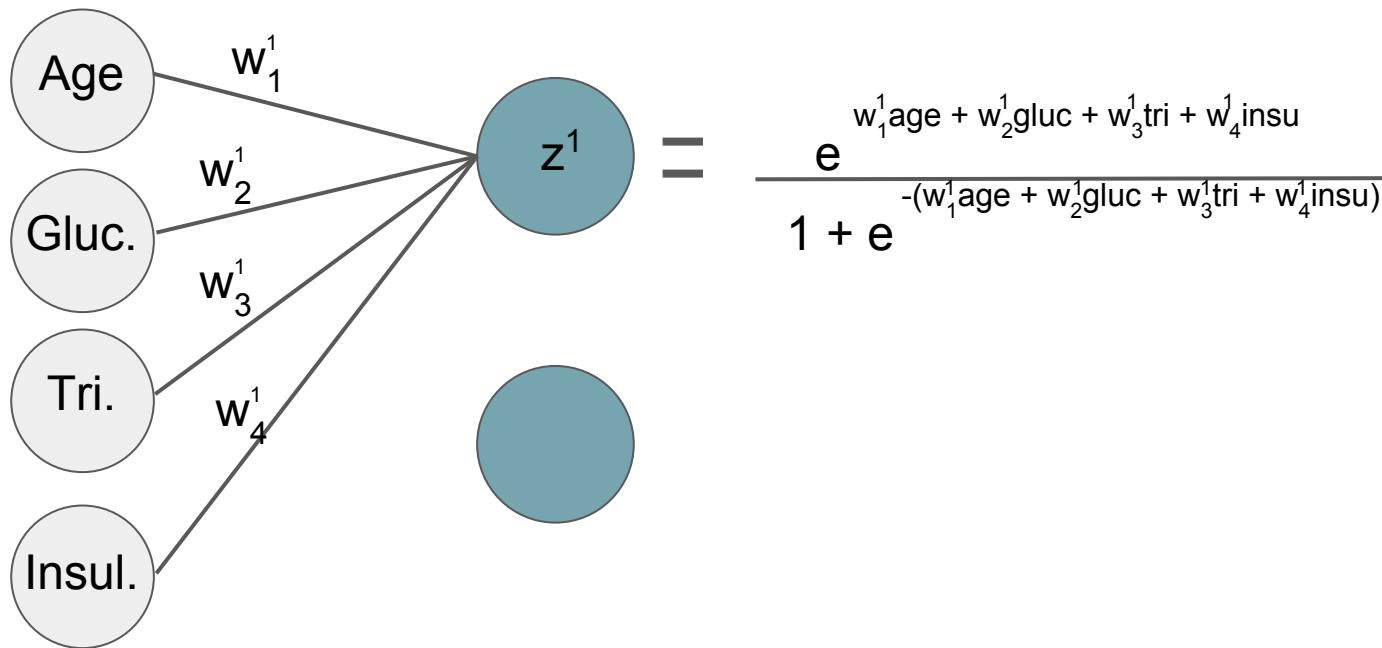
Neural networks (deep learning)



This is called **logistic regression**

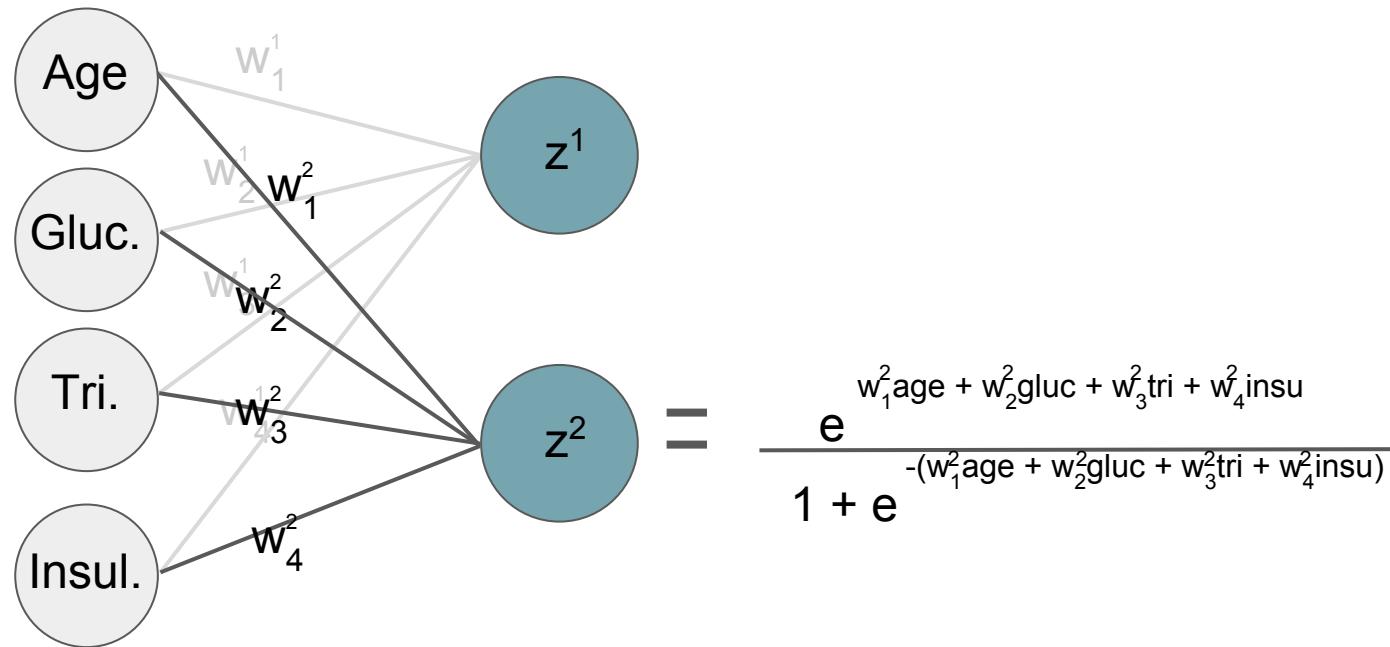
Input features

Neural networks (deep learning)



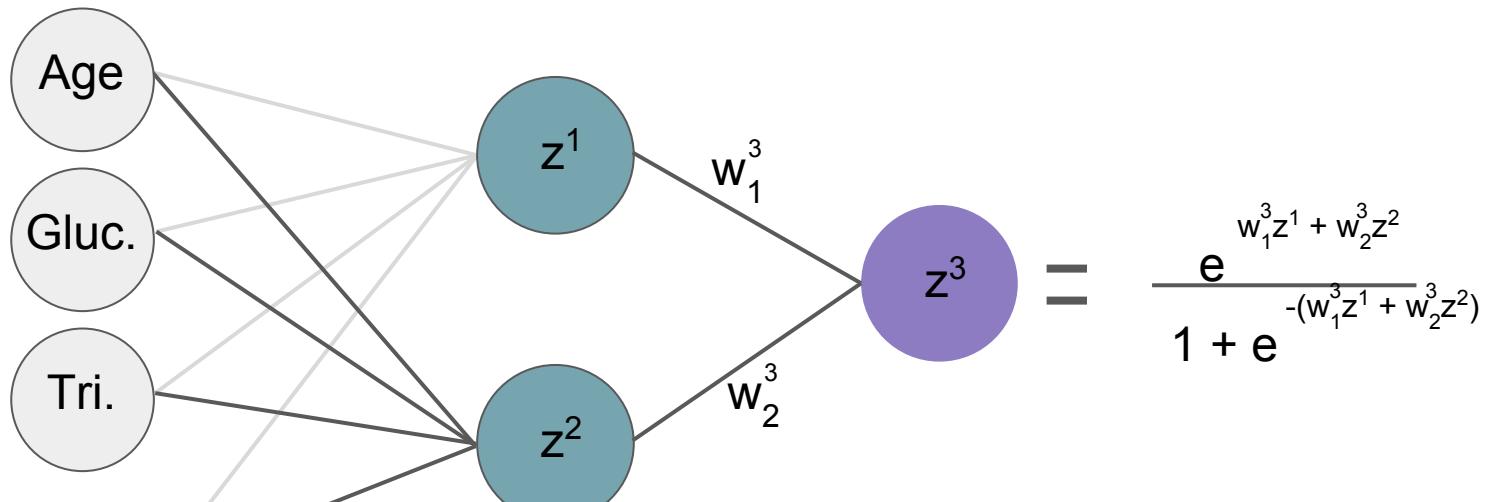
Input features

Neural networks (deep learning)



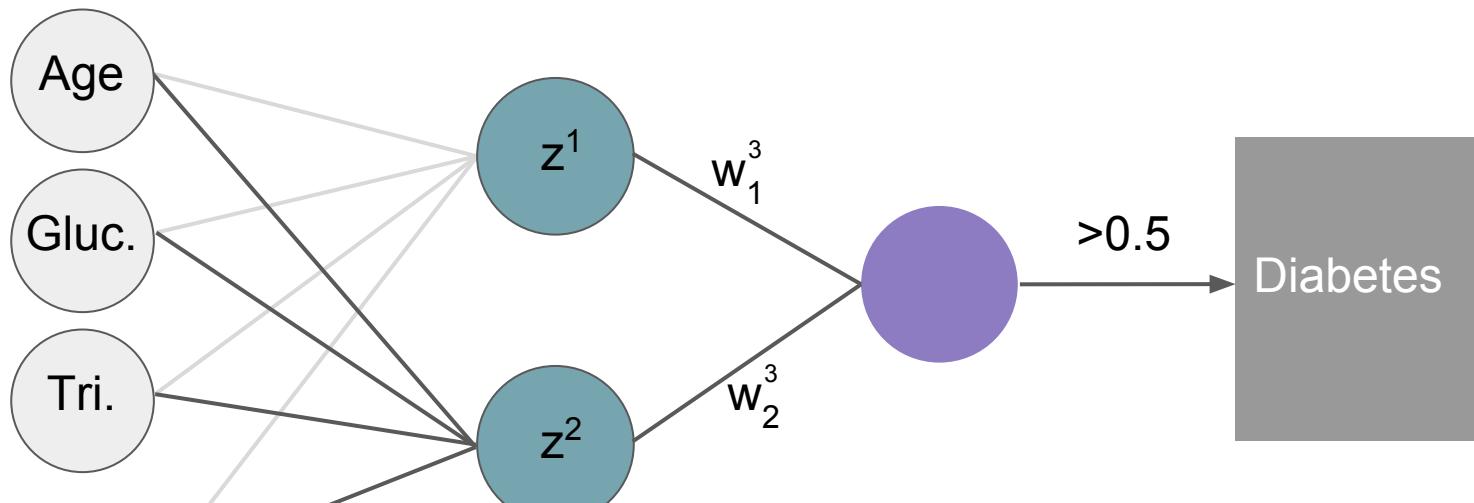
Input features

Neural networks (deep learning)



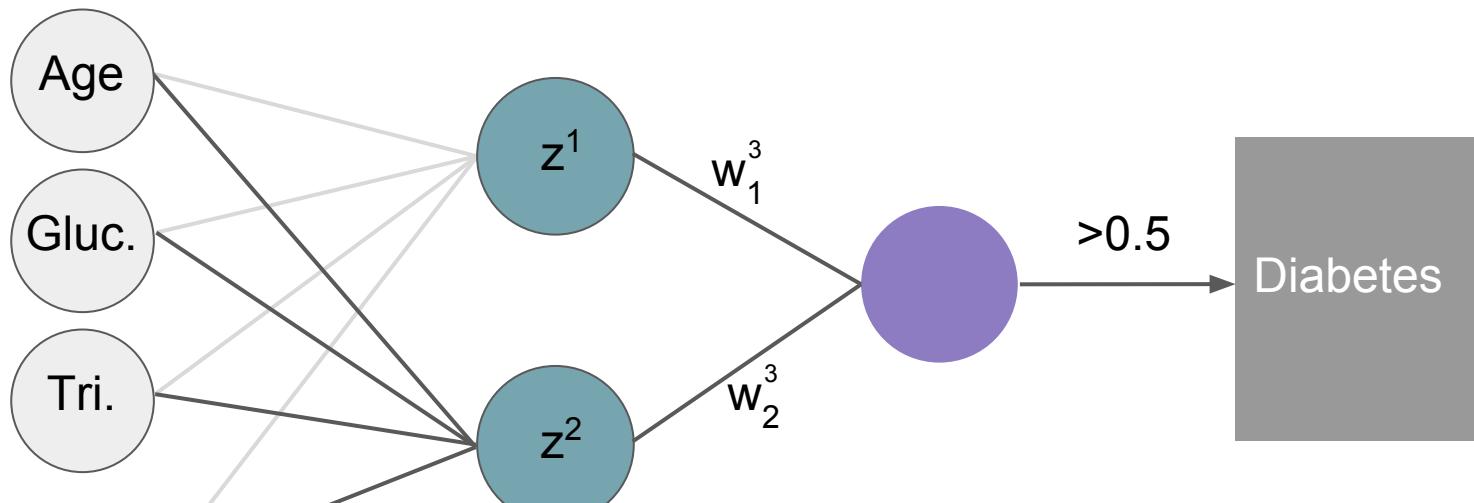
Input features

Neural networks (deep learning)



Input features

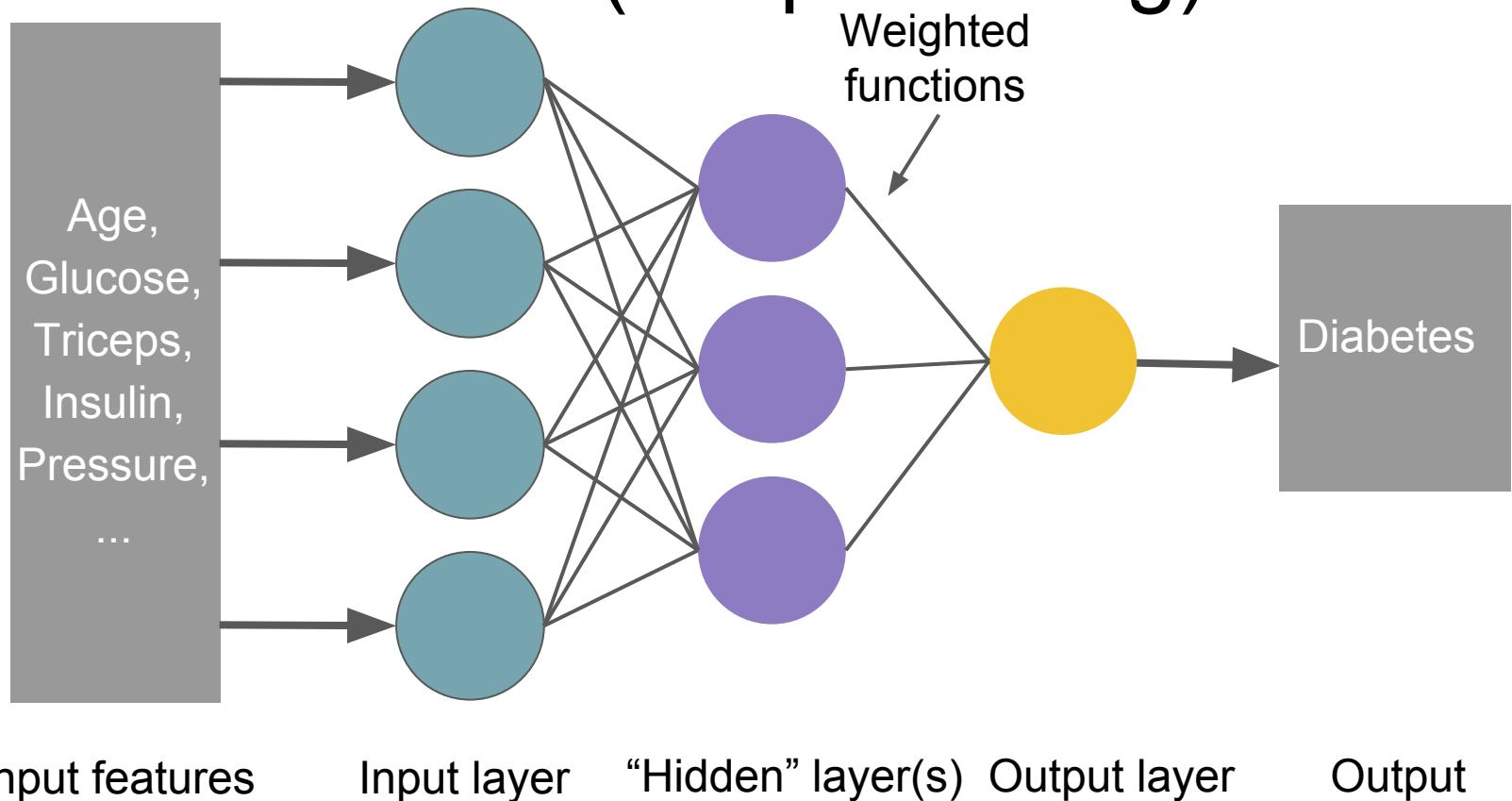
Neural networks (deep learning)



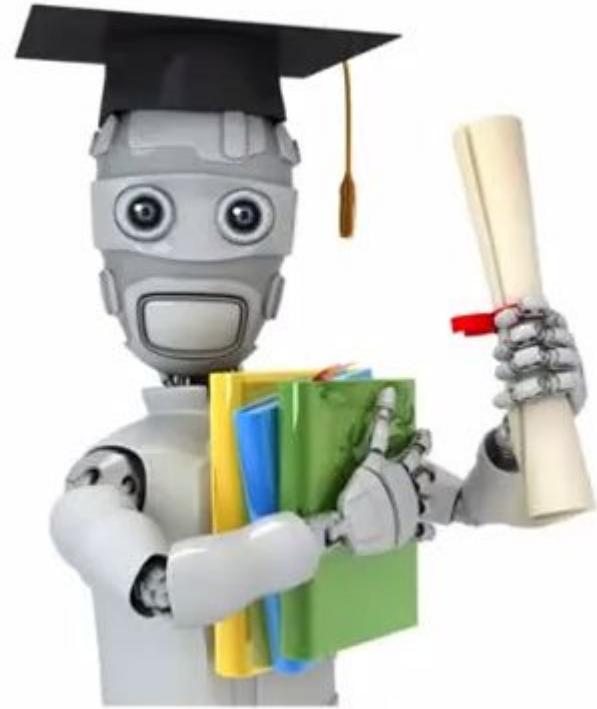
The hard part is “training” the weights, w_i^j .

Input features

Neural networks (deep learning)



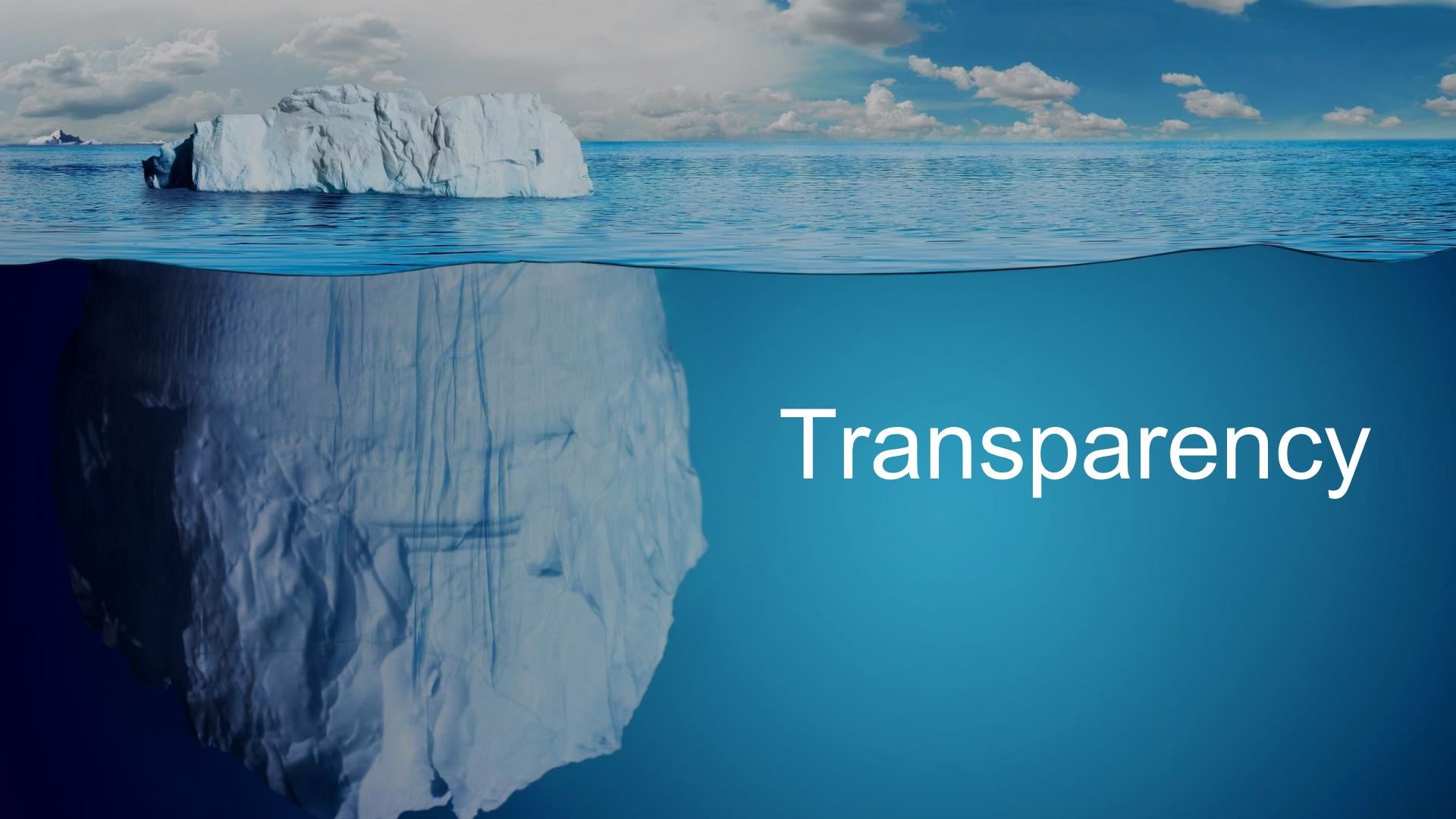
What makes a ML algorithm good?



Machine Learning

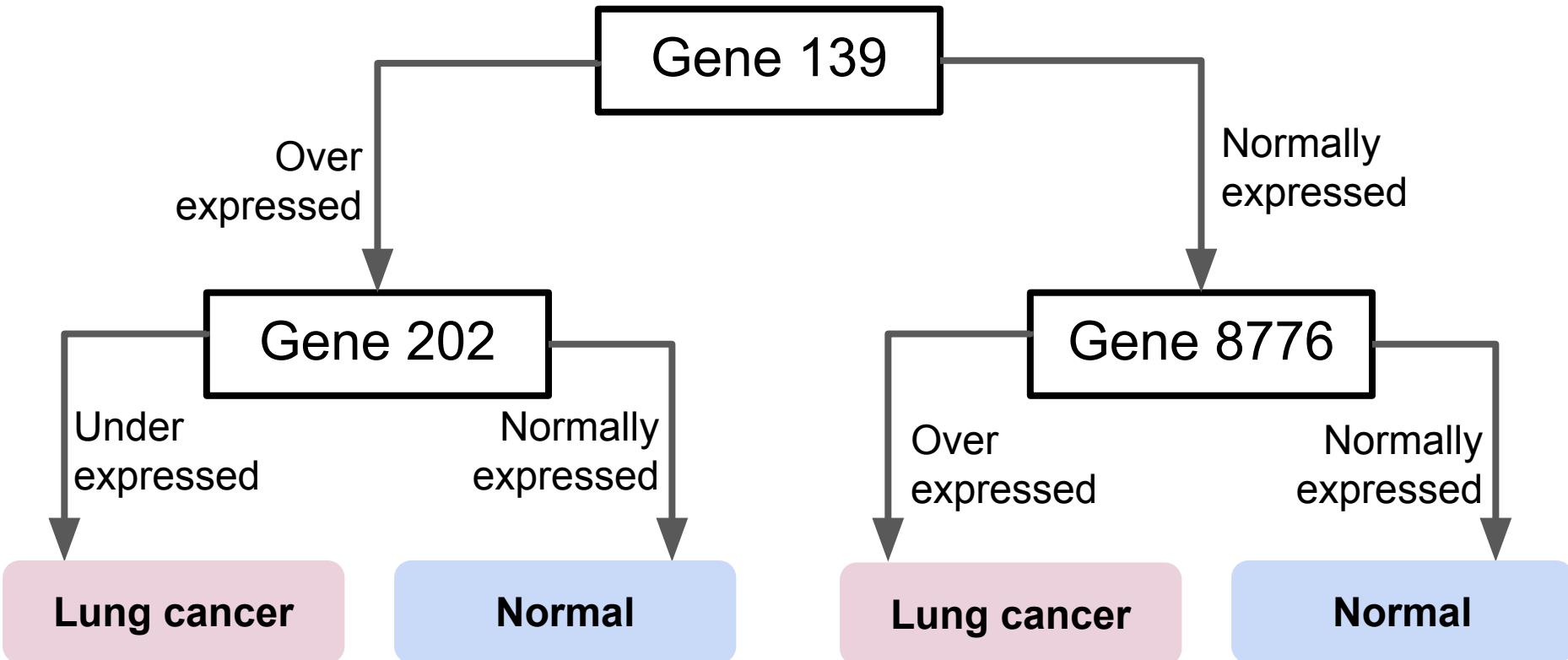
Accuracy



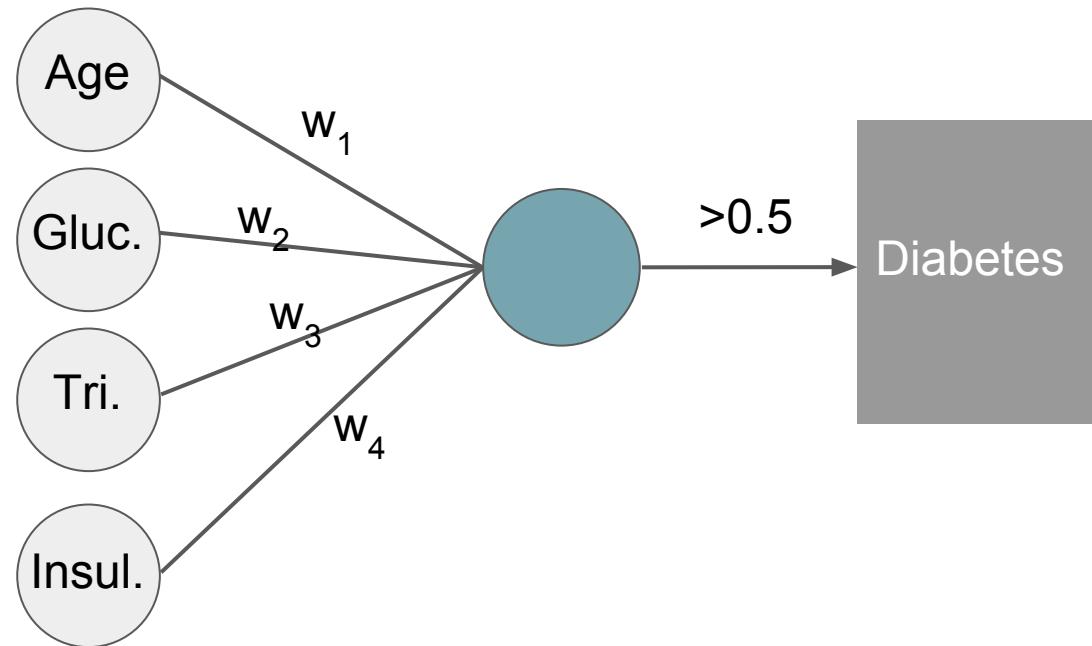
A large, white iceberg is shown floating in a blue ocean under a cloudy sky. The image is split horizontally, showing the massive, textured mass of the iceberg submerged below the water's surface on the left, and its smaller, visible portion above the surface on the right.

Transparency

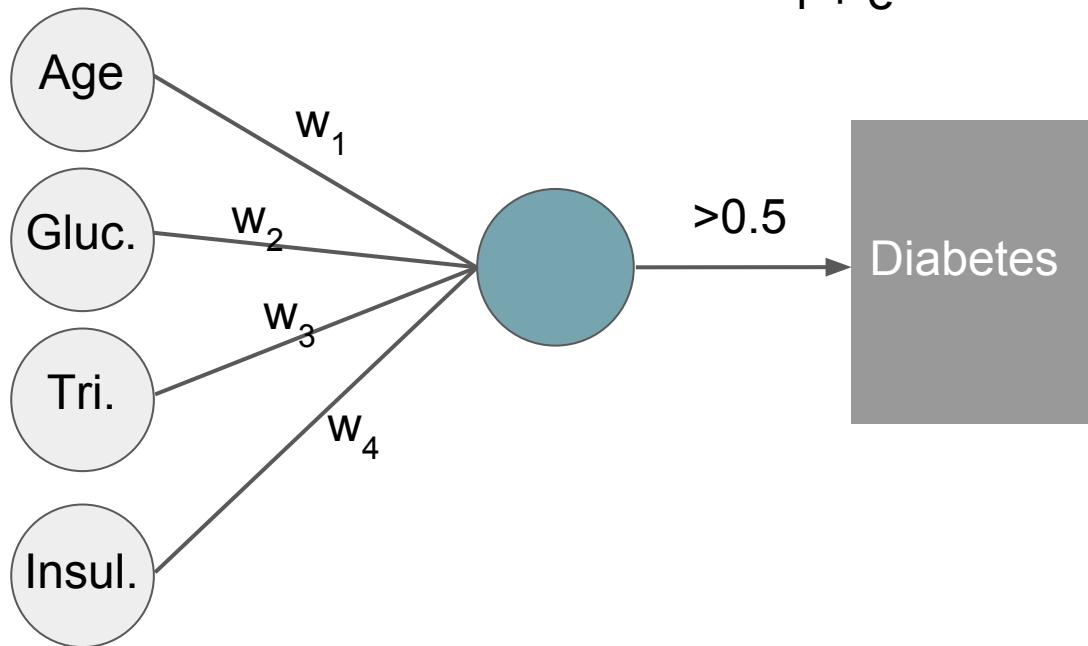
Decision trees



Logistic regression

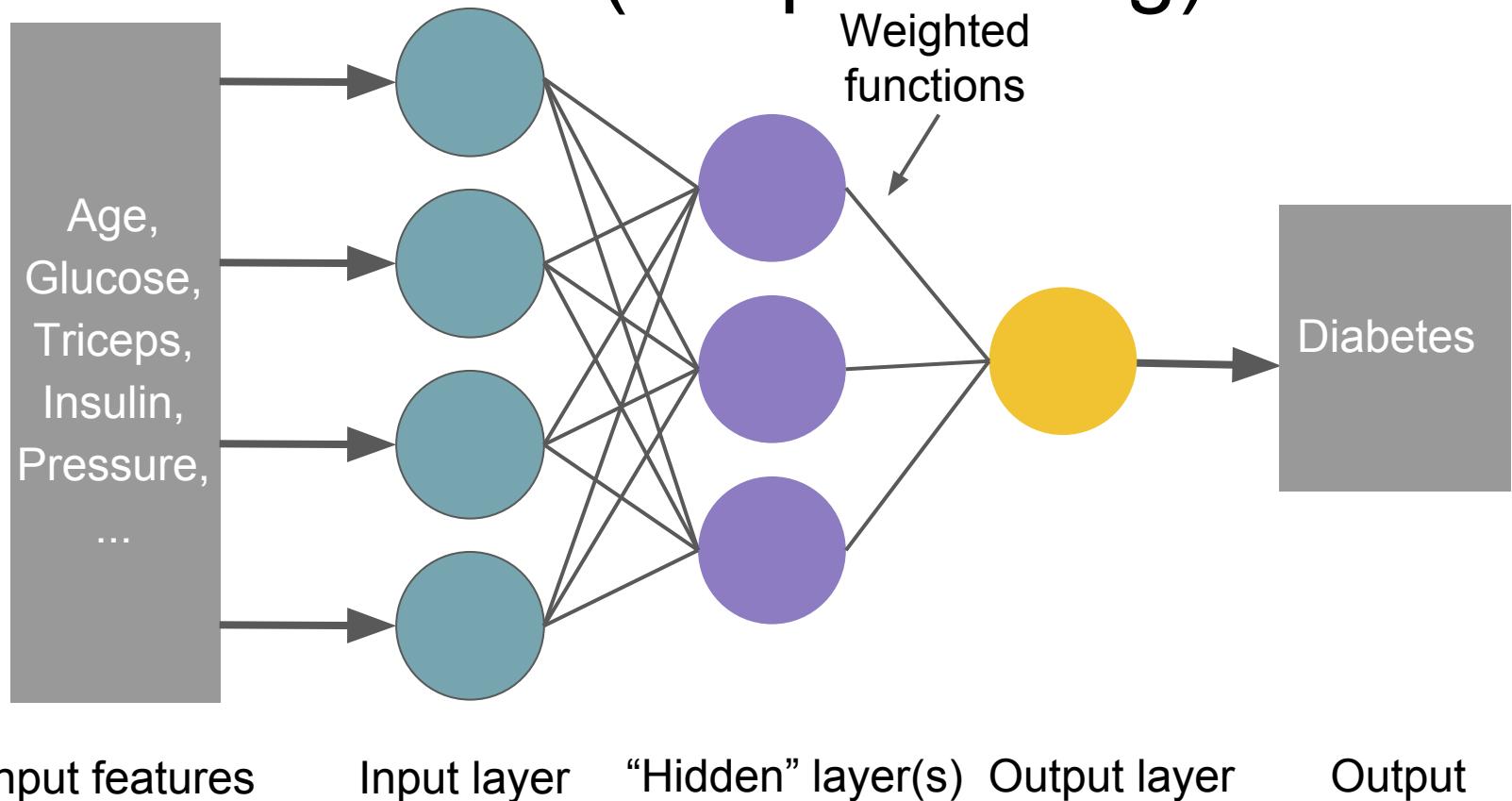


Logistic regression



$$\frac{e^{0.5 \text{ age} + 7.2 \text{ gluc} + 1.2 \text{ tri} + 9.8 \text{ insu}}}{1 + e^{-0.5 \text{ age} - 7.2 \text{ gluc} - 1.2 \text{ tri} - 9.8 \text{ insu}}}$$

Neural networks (deep learning)



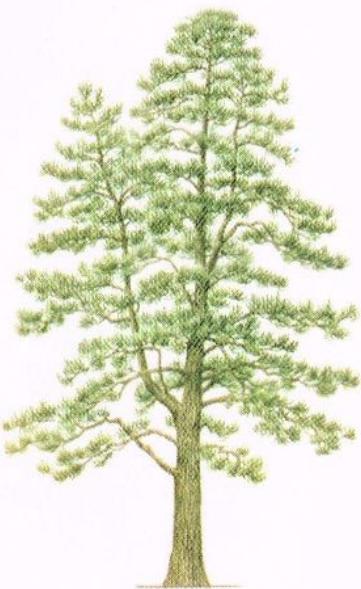
Stability



Computational speed



Generalizability



Black Pine



Pine



Yew



Californian Cypress



Larix

Adaptability: keep learning



How Predictive Analytics Can Help Prevent Infection

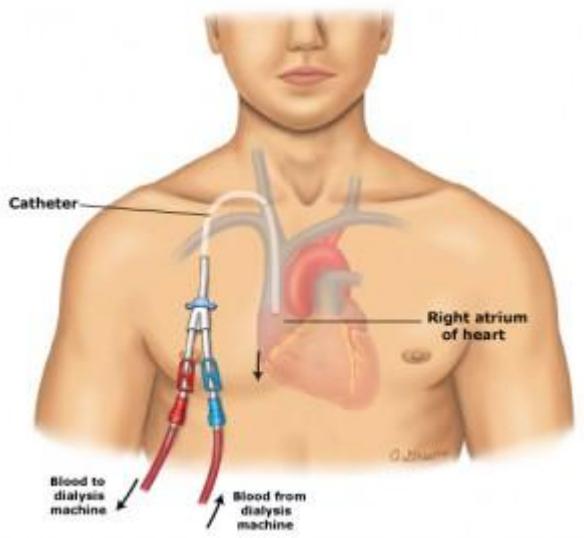
High-powered data tools allow hospitals and systems like Indiana University Health to go beyond a reactive approach

February 7, 2017 By [Rebecca Vesely](#) 0 Comments



Central lines are a fact of life across the better part of Indiana University Health University Hospital as they are in most hospitals. Out of a total of 605,000 inpatient days in 2016, a quarter of those days included a central line. That's not a difficult concept to understand.

But what executives at IU Health, Indianapolis, couldn't accept as a fact of life is that



Central lines and bloodstream infections

A photograph of a hospital ward. In the foreground, an older woman with glasses and a pink shirt is lying in a hospital bed, looking towards the camera. A female nurse in dark blue scrubs is standing beside her, holding a small object. In the background, another patient is lying in a bed, and another nurse is standing nearby, looking at a tablet device. The beds are white with blue accents, and there are medical equipment and curtains in the background.

Goal: predict which patients are most likely to develop such an infection

Enables clinicians to intervene earlier

Predictors

- How long central line has been in a patient

- Length of hospitalization

- Low white blood cell count

- Use of total parenteral nutrition