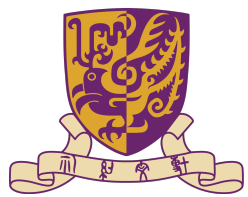
CSCI3230 (ESTR3108)
Fundamentals of Artificial Intelligence

Tutorial 11. AI Applications in Healthcare

Wenao Ma

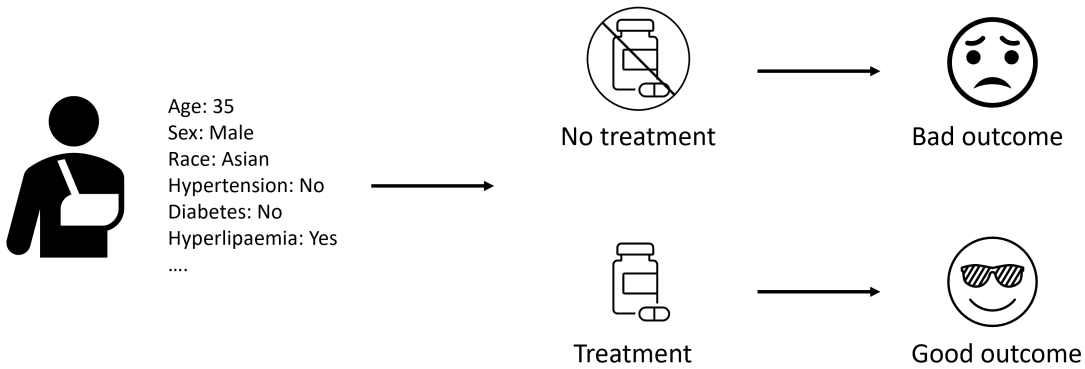
Email: wama@cse.cuhk.edu.hk
Office: Room 1024, 10/F, SHB

Dept. of Computer Science & Engineering
The Chinese University of Hong Kong



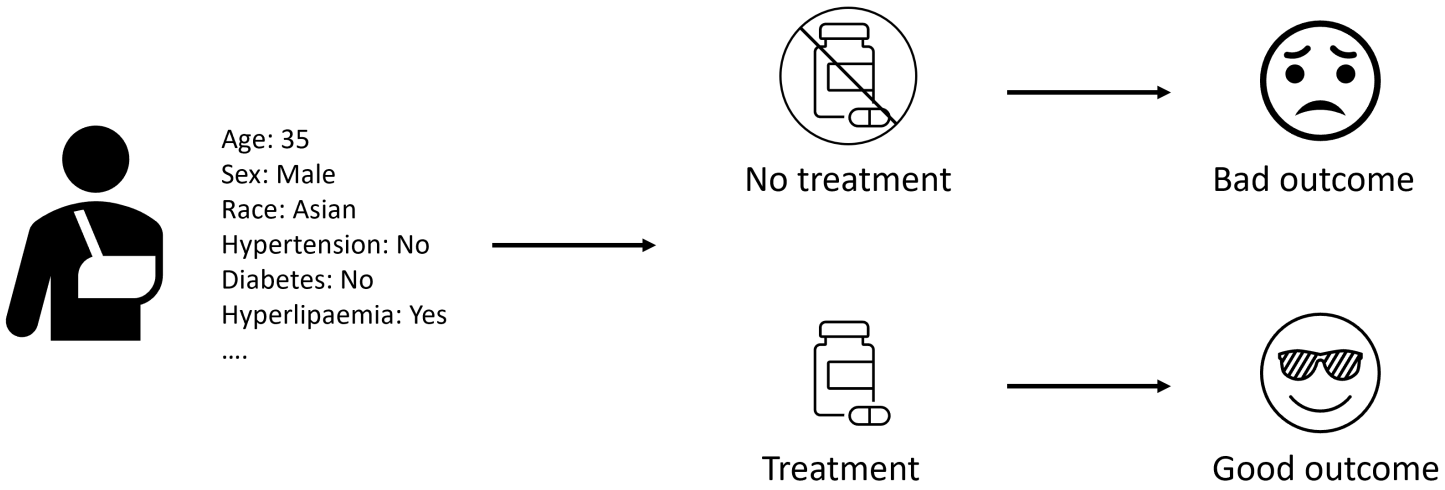
AI Applications in Healthcare

- Today let's briefly introduce an example for AI application in healthcare, and by the way, review the contents we have learned in this course.
- I will **highlight** some basic concepts you have learned from this course. So that you can see the usefulness of these algorithms in practice.



A example: treatment effect estimation

- Task: To predict the treatment effect given the clinical information of a patient.

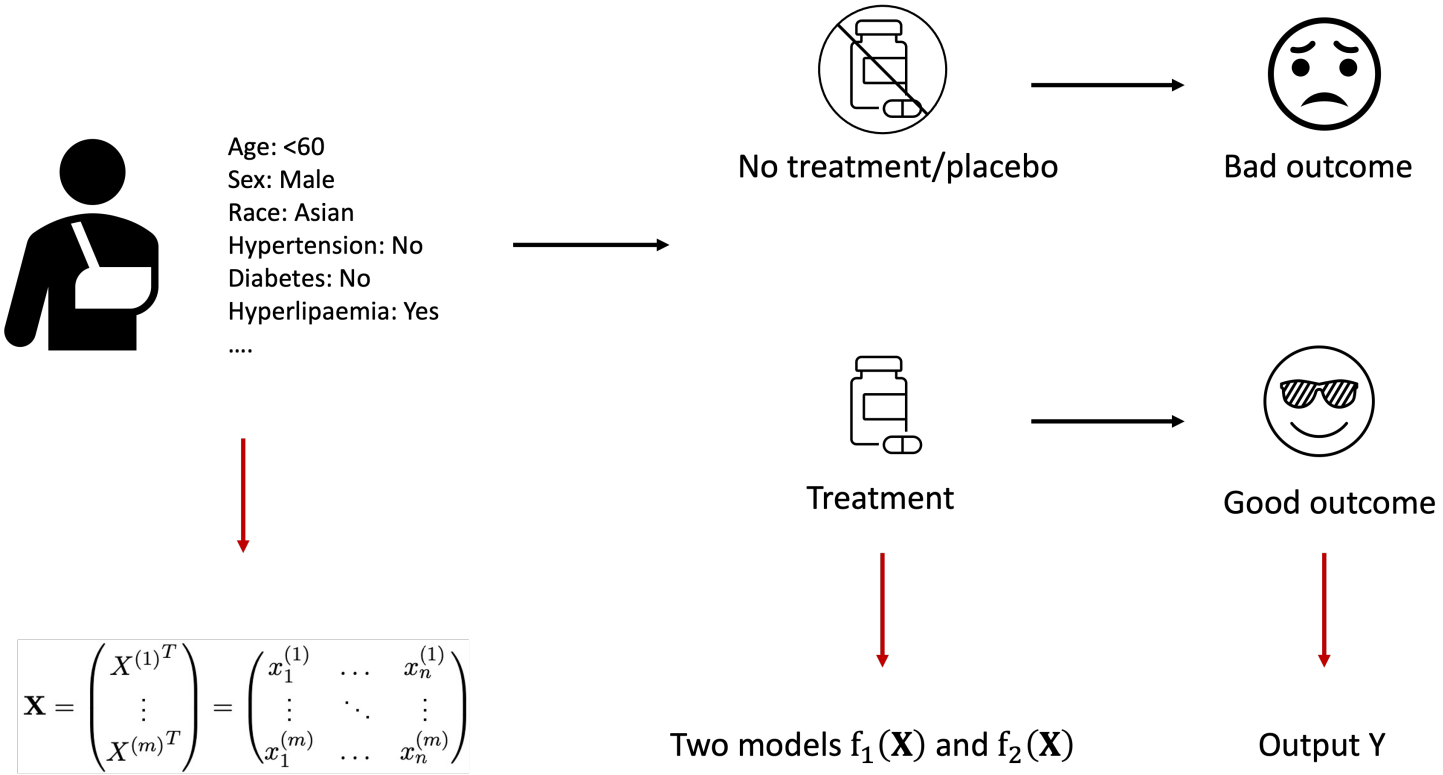


A example: treatment effect estimation

- Let's use machine learning algorithm to handle this problem.

A example: treatment effect estimation

- Let's use machine learning algorithm to handle this problem.



A example: treatment effect estimation

- If the outcomes are continuous variables (e.g., survival time: 1 year, 2 years, 3 years ...), then we can consider it as a regression problem.

Output Y:

Survival time: 1 year

Survival time: 30 year

A example: treatment effect estimation

- If the outcomes are continuous variables (e.g., survival time: 1 year, 2 years, 3 years ...), then we can consider it as a regression problem.

Output Y:

Survival time: 1 year

Survival time: 30 year

- As we have learned before, we can use **Linear Regression Model** to handle this problem:

Multivariate Linear Function

Write the multivariate linear function in matrix form:

$$\hat{y} = \hat{f}_{\Theta,\theta_0}(X) = X^T\Theta + \theta_0$$

where $\Theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}$, and $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

A example: treatment effect estimation

- If the outcomes are discrete variables (e.g., whether the patient will be survival or dead in one year), then we can consider it as a classification problem.

A example: treatment effect estimation

Output Y:

Dead in one year

Survival in one year

- If the outcomes are discrete variables (e.g., whether the patient will be survival or dead in one year), then we can consider it as a classification problem.
- We have learned many algorithms that can be used as a classification model, including **1) Logistic Regression, 2) Support Vector Machine, 3) Clustering Algorithms and 4) Neural Network.**

A example: treatment effect estimation

- If the outcomes are discrete variables (e.g., whether the patient will be survival or dead in one year), then we can consider it as a classification problem.
- We have learned many algorithms that can be used as a classification model, including **1) Logistic Regression, 2) Support Vector Machine, 3) Clustering Algorithms and 4) Neural Network.**
- If the training data have no labels of outcomes, we can use **Clustering Algorithms** as an **Unsupervised Learning** method. The remaining classification models usually use labels to perform **Supervised Learning**, if the labels of outcomes are avaiable.

A example: treatment effect estimation

- **Logistic Regression** (apply logit transformation on the linear function)

$$\boxed{P(\hat{y} = 1|X) = \text{logit}^{-1}(X^T\Theta + \theta_0) = \frac{1}{1 + e^{-(X^T\Theta + \theta_0)}}}$$

A example: treatment effect estimation

- **Logistic Regression** (apply logit transformation on the linear function)

$$\boxed{P(\hat{y} = 1|X) = \text{logit}^{-1}(X^T\Theta + \theta_0) = \frac{1}{1 + e^{-(X^T\Theta + \theta_0)}}}$$

- **Support Vector Machine** (Another linear classifer model which uses different method to optimize the model)

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 \quad \text{s.t.} \quad y_i\left(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x_i} + b\right) \geq 1$$

A example: treatment effect estimation

- **Logistic Regression** (apply logit transformation on the linear function)

$$\boxed{P(\hat{y} = 1|X) = \text{logit}^{-1}(X^T\Theta + \theta_0) = \frac{1}{1 + e^{-(X^T\Theta + \theta_0)}}}$$

- **Support Vector Machine** (Another linear classifer model which uses different method to optimize the model)

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 \quad \text{s.t.} \quad y_i\left(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x_i} + b\right) \geq 1$$

- **Clustering** (three popular algorithms: 1). K-Means, 2). DBSCAN and 3). Agglomerative hierarchical clustering)

A example: treatment effect estimation

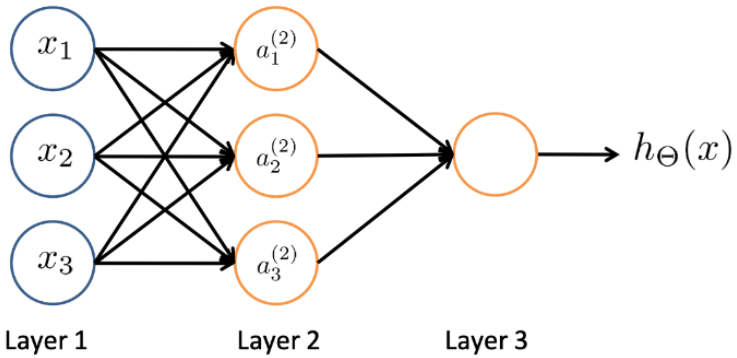
- **Logistic Regression** (apply logit transformation on the linear function)

$$P(\hat{y} = 1|X) = \text{logit}^{-1}(X^T\Theta + \theta_0) = \frac{1}{1 + e^{-(X^T\Theta+\theta_0)}}$$

- **Support Vector Machine** (Another linear classifer model which uses different method to optimize the model)

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 \quad \text{s.t.} \quad y_i\left(\boldsymbol{w}^T\boldsymbol{x_i} + b\right) \geq 1$$

- **Clustering** (three popular algorithms: 1). K-Means, 2). DBSCAN and 3). Agglomerative hierarchical clustering)
- **Neural Network:**



A example: treatment effect estimation

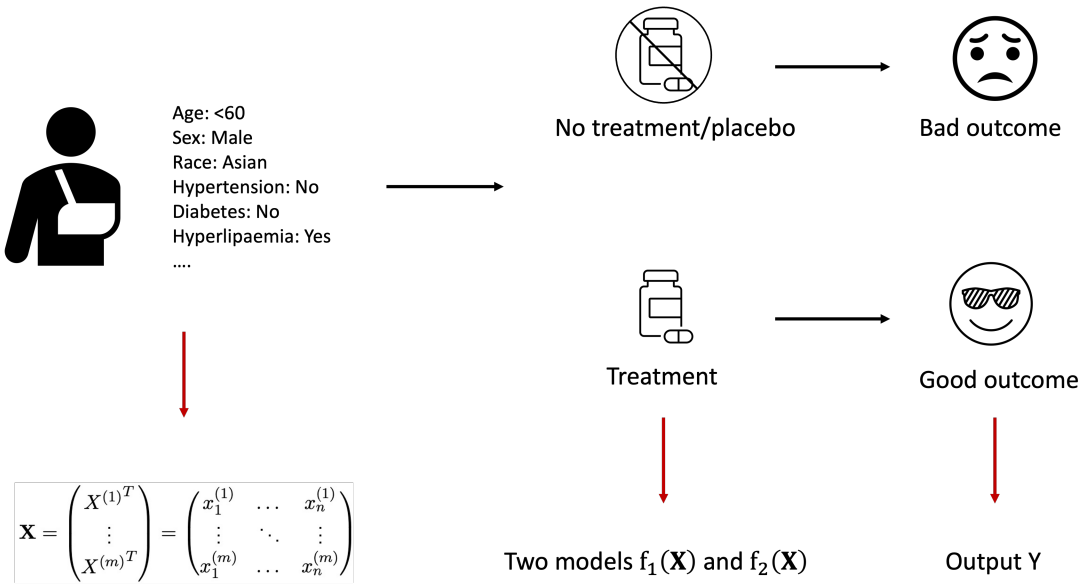
- After training the regression model or classification model, you can use test data and some **Evaluation Metrics** to validate the model performance in real world.

A example: treatment effect estimation

- After training the regression model or classification model, you can use test data and some **Evaluation Metrics** to validate the model performance in real world.
- And you may be faced with **Underfitting** or **Overfitting** problem, which are related to the number of data and complexity of model

A example: treatment effect estimation

- After training the regression model or classification model, you can use test data and some **Evaluation Metrics** to validate the model performance in real world.
- And you may be faced with **Underfitting** or **Overfitting** problem, which are related to the number of data and complexity of model

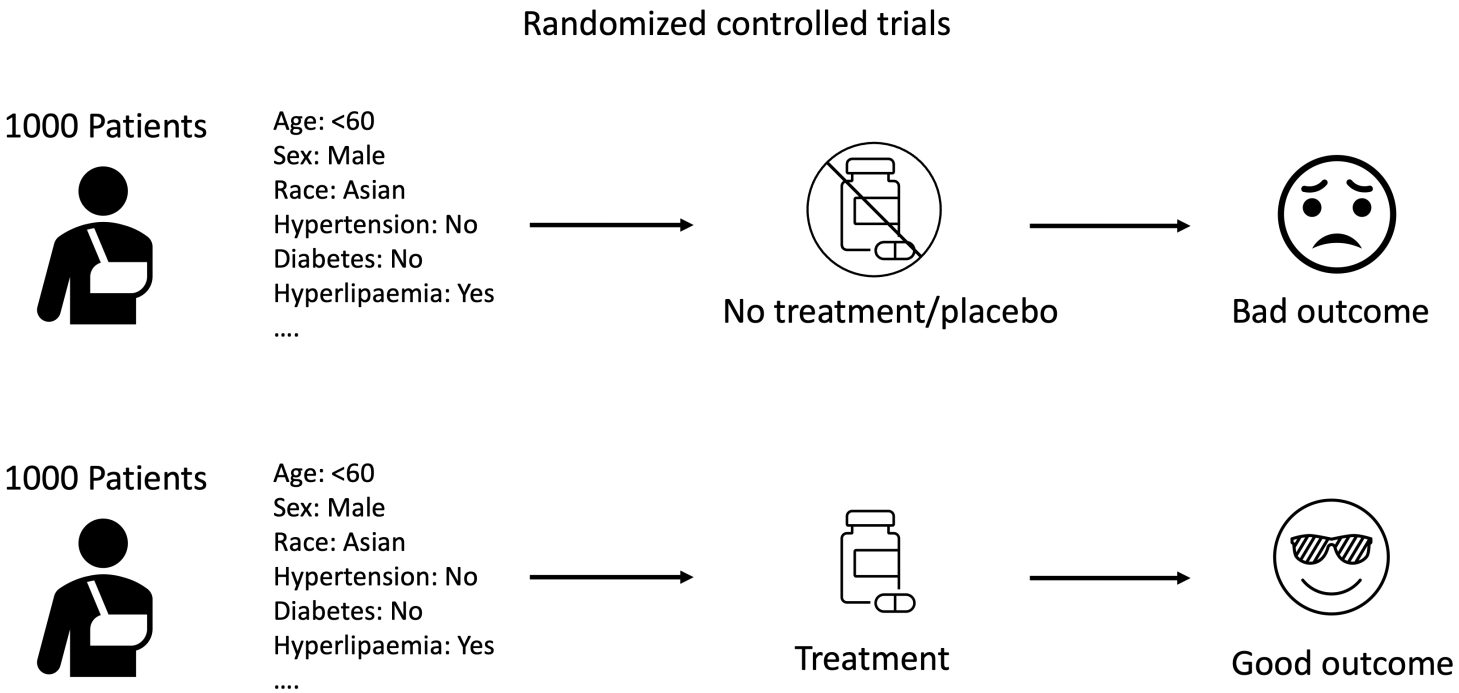


A example: treatment effect estimation

- Let's go back to the task.

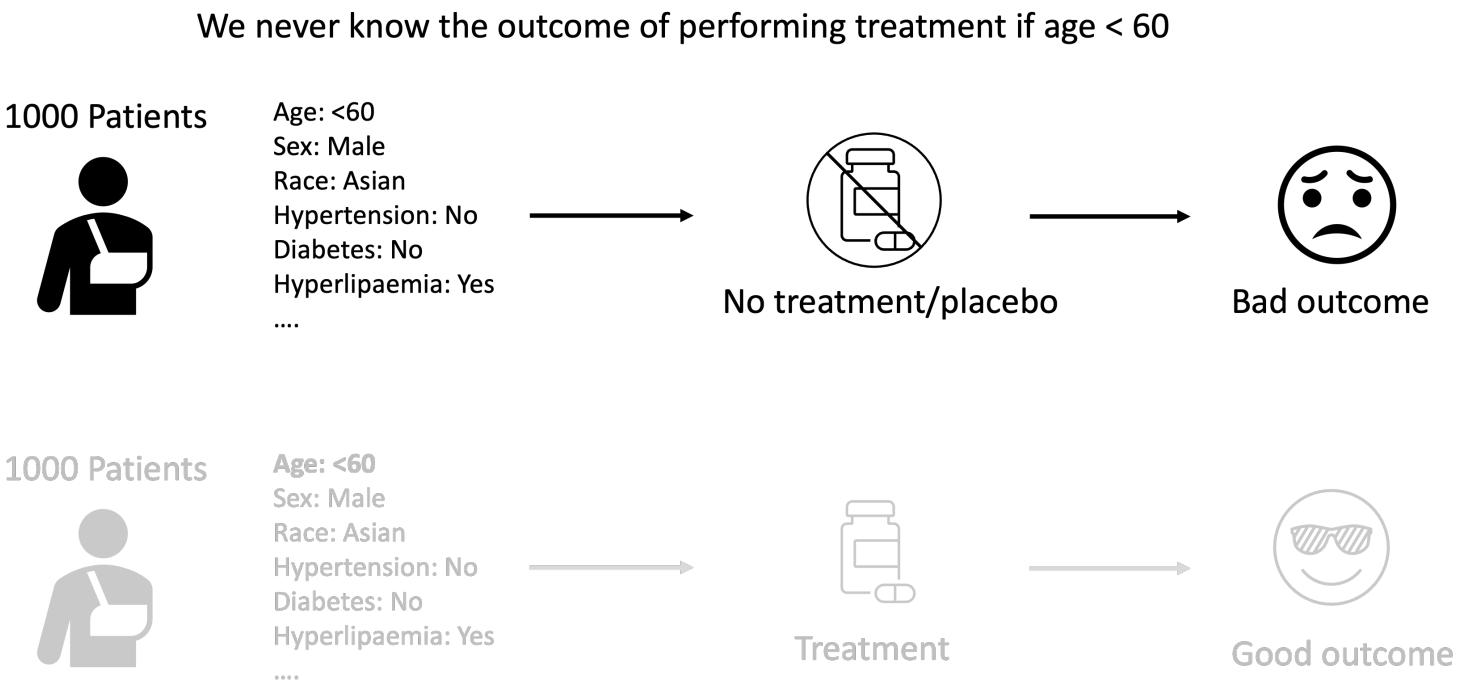
A example: treatment effect estimation

- Let's go back to the task.
- In clinical scenarios, we conduct randomized controlled trials to test the outcome of each treatment strategy.



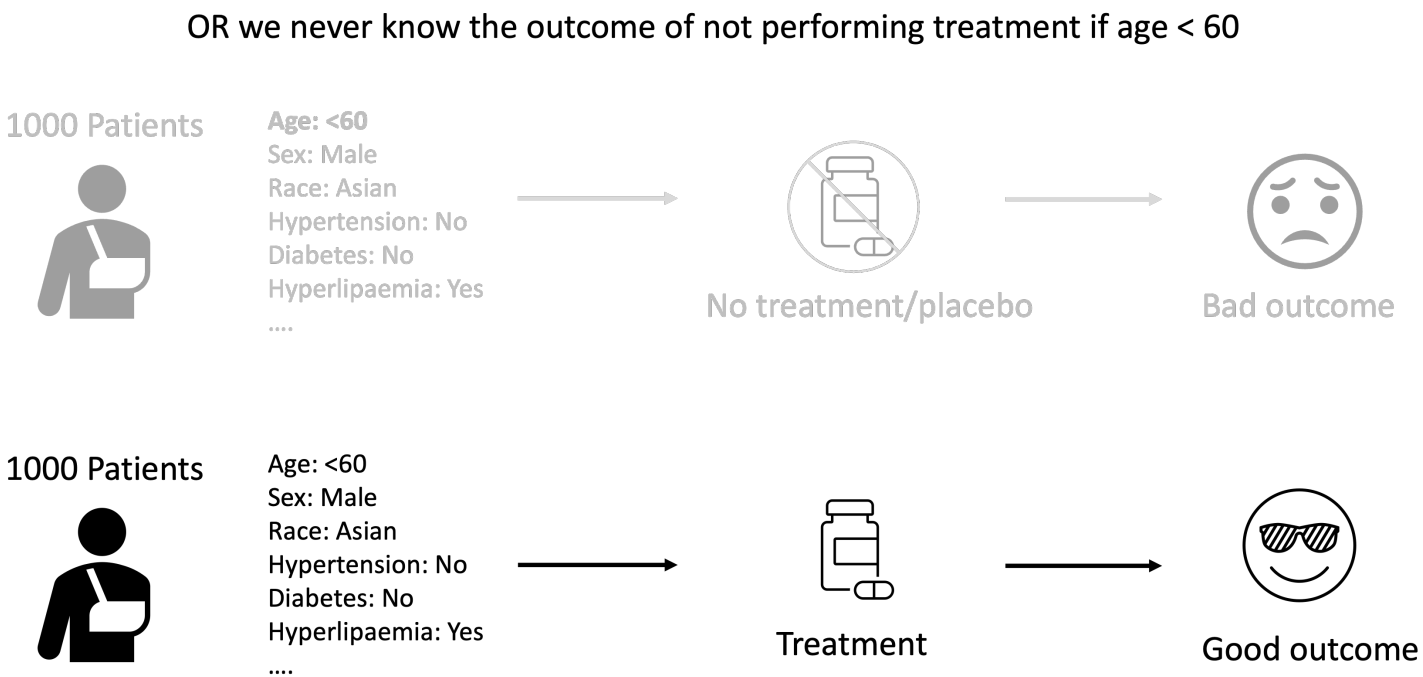
A example: treatment effect estimation

- However, sometimes it's hard for us to obtain the data of randomized controlled trials. In many clinical scenarios, the treatment assignment is biased.



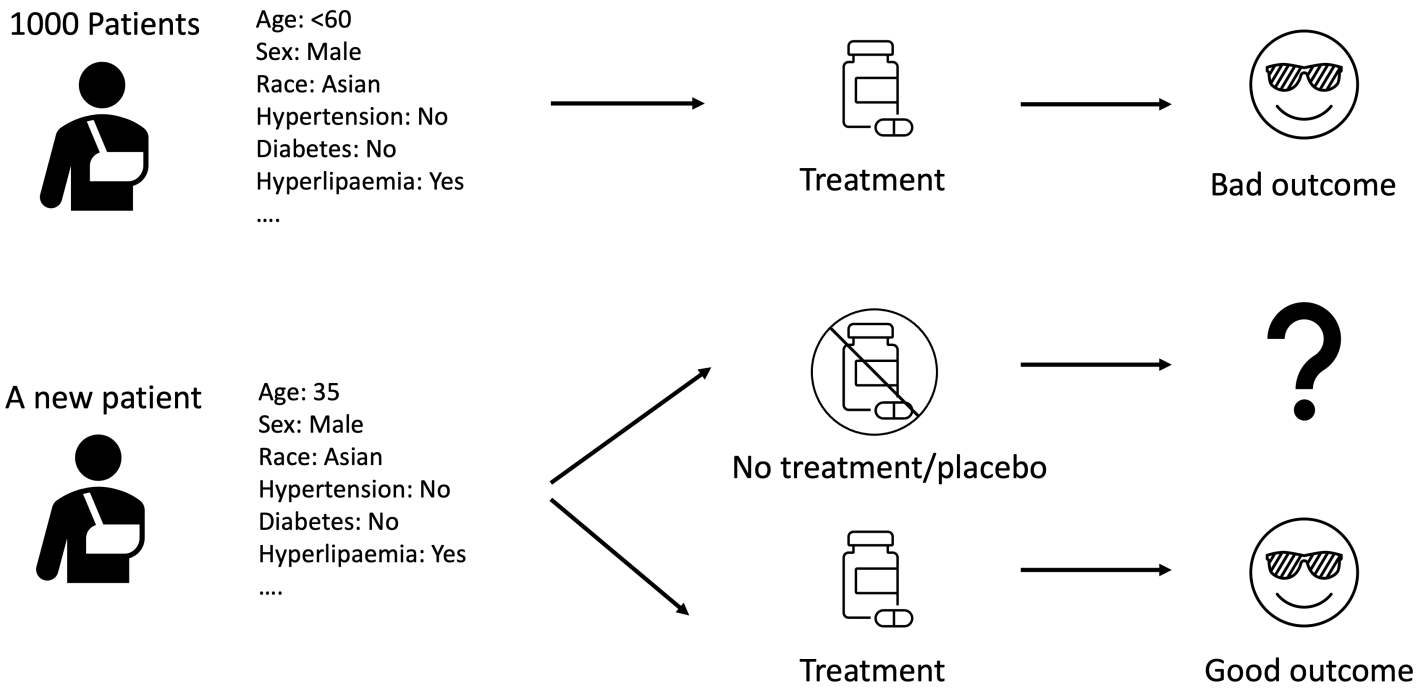
A example: treatment effect estimation

- However, sometimes it's hard for us to obtain the data of randomized controlled trials. In many clinical scenarios, the treatment assignment is biased.



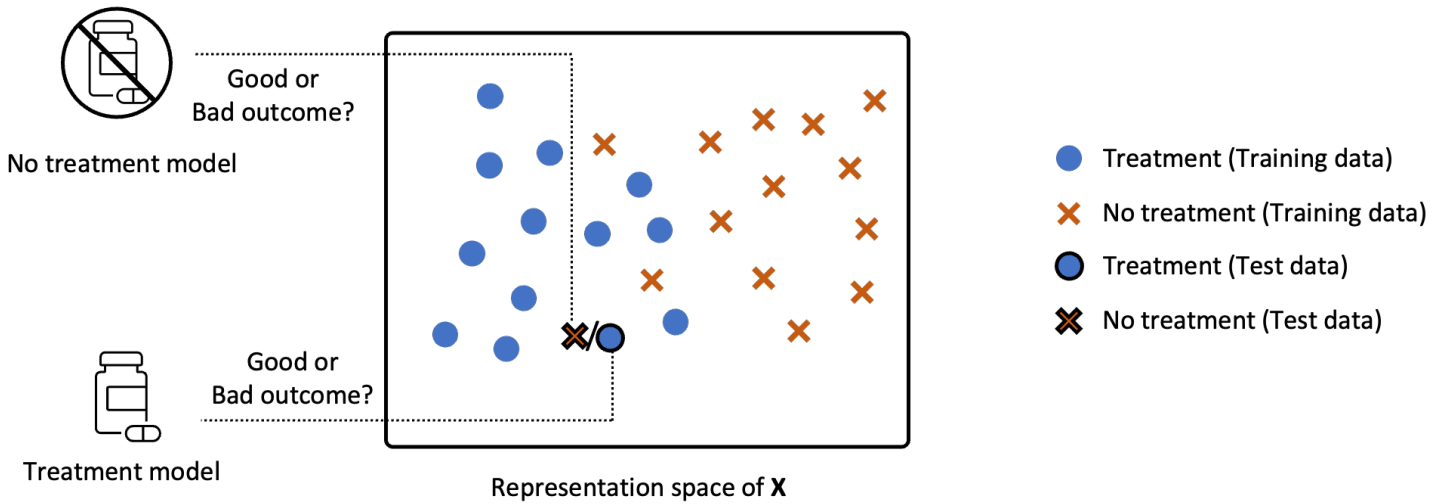
A example: treatment effect estimation

- So it's hard for us to estimate the outcome conditioned on some situations.



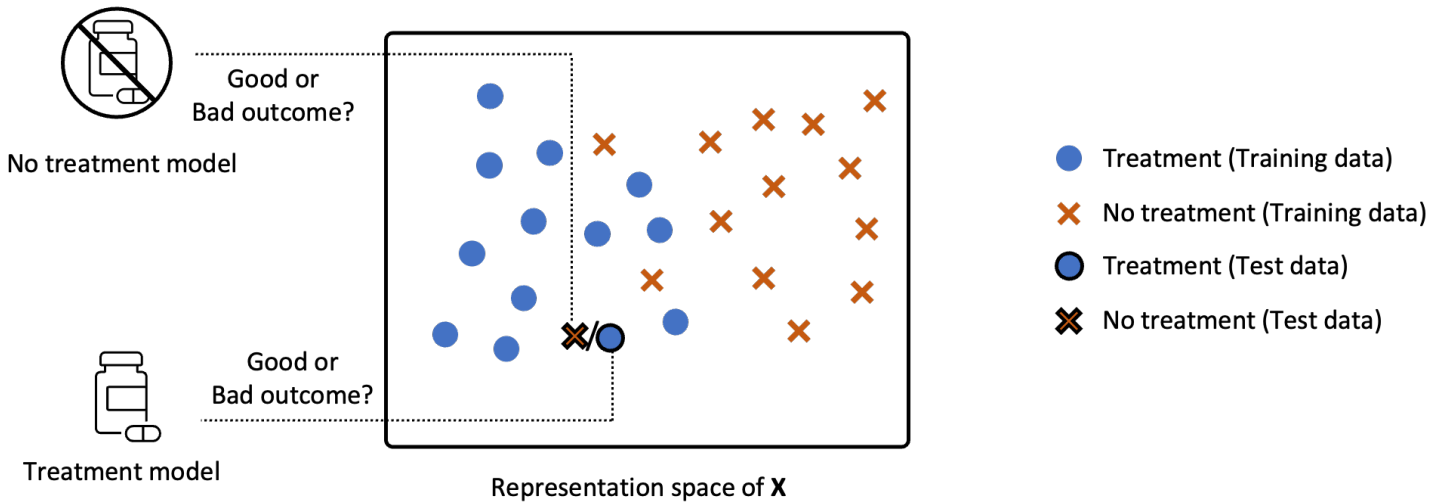
A example: treatment effect estimation

- From the perspective of machine learning, the key point is that, the features of the test data is unseen conditioned on a certain treatment assignment.
- In other words, in the representation space, there are no training data around the test data, which makes the outcome be unconfident.



A example: treatment effect estimation

- From the perspective of machine learning, the key point is that, the features of the test data is unseen conditioned on a certain treatment assignment.
- In other words, in the representation space, there are no training data around the test data, which makes the outcome be unconfident.



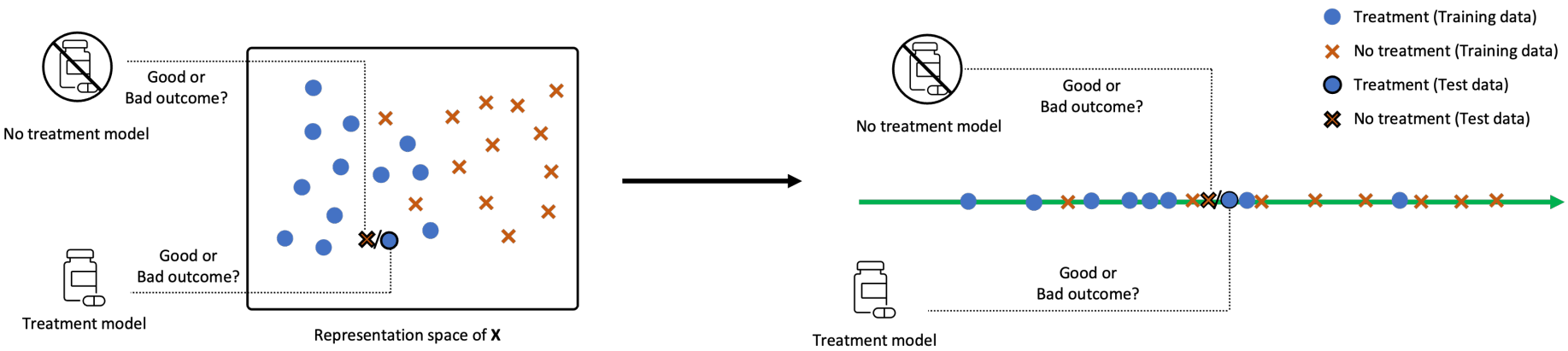
- How to deal with this problem?

A example: treatment effect estimation

- One of the solution is performing dimensionality reduction to increase the overlap.

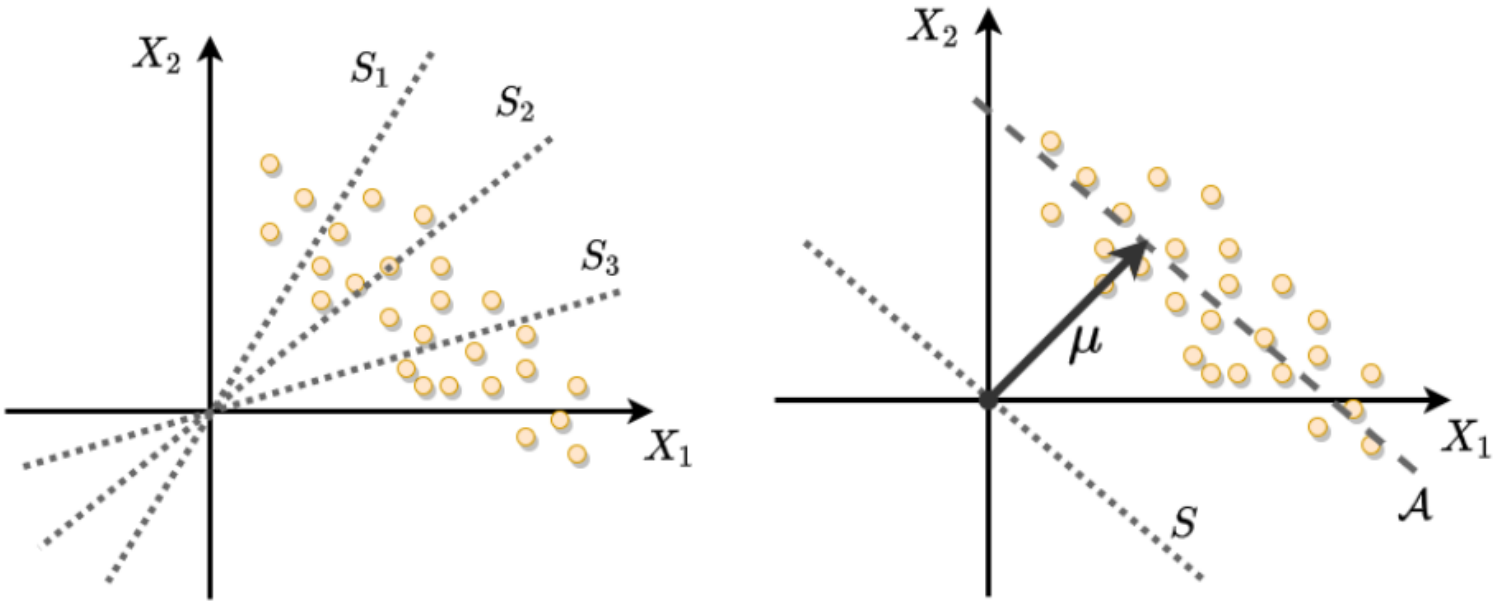
A example: treatment effect estimation

- One of the solution is performing dimensionality reduction to increase the overlap.
- After the dimensionality reduction, if in the representation space, there are more training data around the test data, we can say that we are more confident about the outcome prediction.



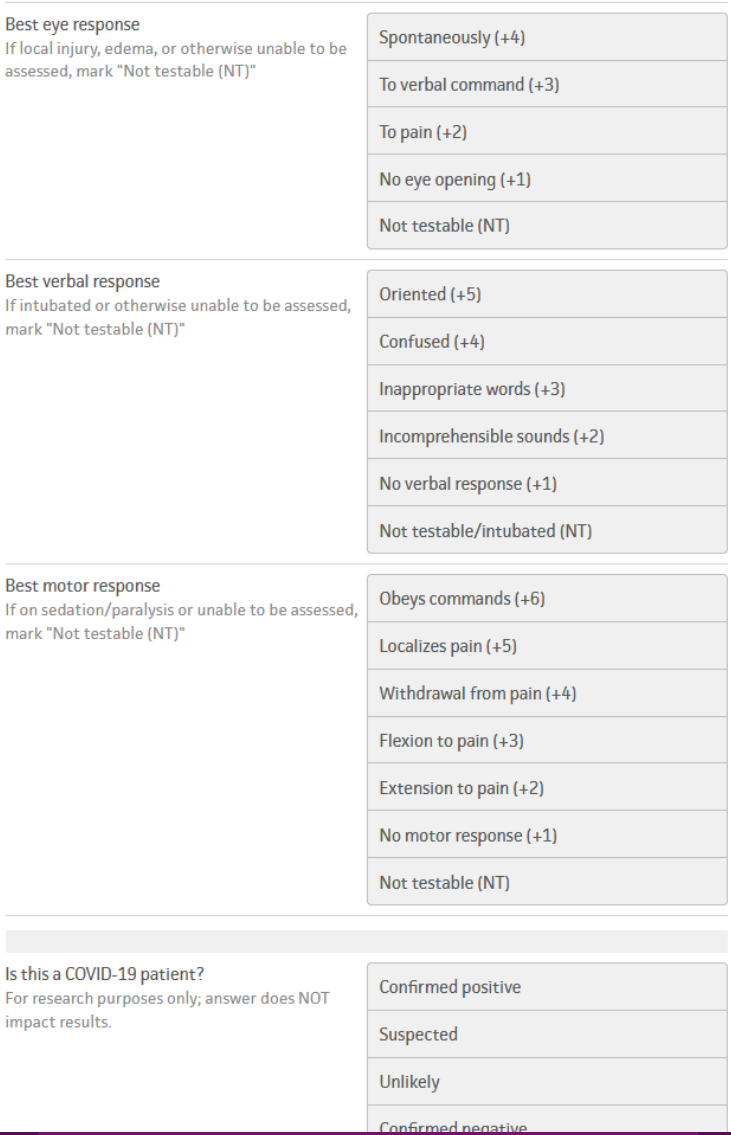
A example: treatment effect estimation

- Recall that **Principal Component Analysis** can be used as a method to perform dimensionality reduction:



A example: treatment effect estimation

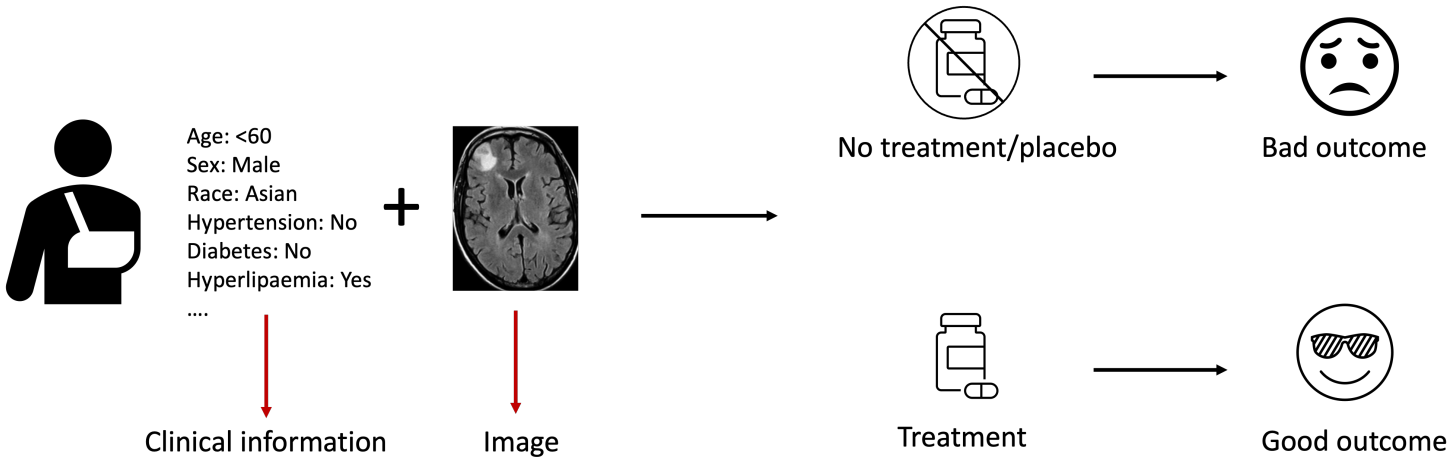
- It's similar to some prognostic scores used in clinical scenarios:



Best eye response If local injury, edema, or otherwise unable to be assessed, mark "Not testable (NT)"	Spontaneously (+4) To verbal command (+3) To pain (+2) No eye opening (+1) Not testable (NT)
Best verbal response If intubated or otherwise unable to be assessed, mark "Not testable (NT)"	Oriented (+5) Confused (+4) Inappropriate words (+3) Incomprehensible sounds (+2) No verbal response (+1) Not testable/intubated (NT)
Best motor response If on sedation/paralysis or unable to be assessed, mark "Not testable (NT)"	Obeys commands (+6) Localizes pain (+5) Withdrawal from pain (+4) Flexion to pain (+3) Extension to pain (+2) No motor response (+1) Not testable (NT)
Is this a COVID-19 patient? For research purposes only; answer does NOT impact results.	Confirmed positive Suspected Unlikely Confirmed negative

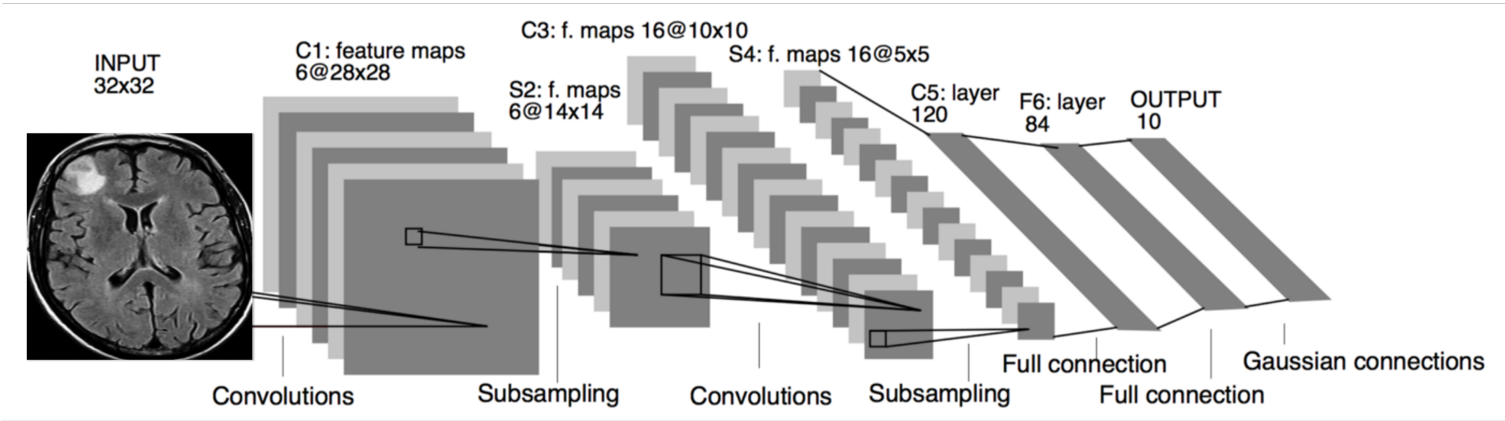
A example: treatment effect estimation

- In some scenarios, we also want to combine both the image information and clinical information to conduct outcome prediction:



A example: treatment effect estimation

- **Convolutional Neural Networks** can be a powerful method to extract feature and perform classification task for the image data.



A example: treatment effect estimation

- Actually we often combine many algorithms in practice, e.g., use convolutional neural networks to extract features and then use clustering algorithms to perform unsupervised learning.

A example: treatment effect estimation

- Actually we often combine many algorithms in practice, e.g., use convolutional neural networks to extract features and then use clustering algorithms to perform unsupervised learning.
- You should also pay attention to the details of each algorithm (from Lecture 1 to Lecture 10) to prepare the final exam.

A example: treatment effect estimation

- Actually we often combine many algorithms in practice, e.g., use convolutional neural networks to extract features and then use clustering algorithms to perform unsupervised learning.
- You should also pay attention to the details of each algorithm (from Lecture 1 to Lecture 10) to prepare the final exam.
- This example has not mentioned the algorithms of **Uniformed Search** and **Informed Search**. You can review them in Lecture 9 and Lecture 10.