

Моделі статистичного навчання: вступ

Навчання з учителем та навчання без учителя.

Навчання з учителем та навчання без учителя.

Навчання з учителем: побудова моделі для передбачення чи оцінки результату на основі вхідних даних.

Навчання з учителем та навчання без учителя.

Навчання з учителем: побудова моделі для передбачення чи оцінки результату на основі вхідних даних.

Навчання без учителя: наявними є вхідні дані, але відсутні відповідні результати.

Навчання з учителем та навчання без учителя.

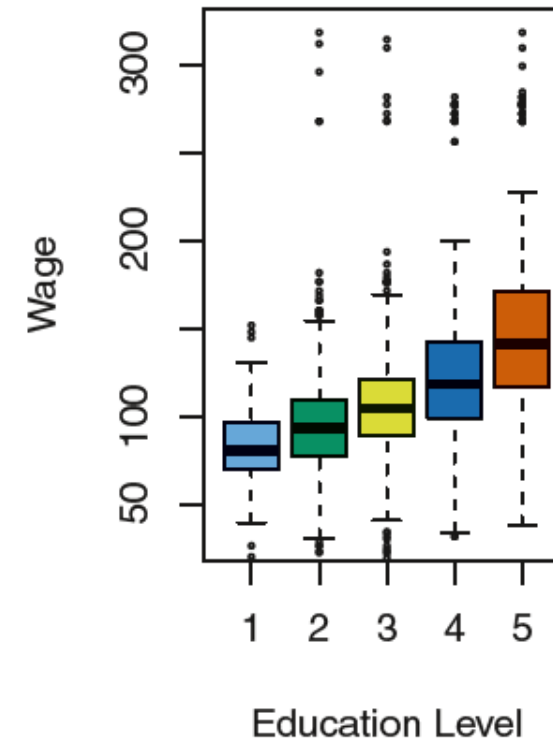
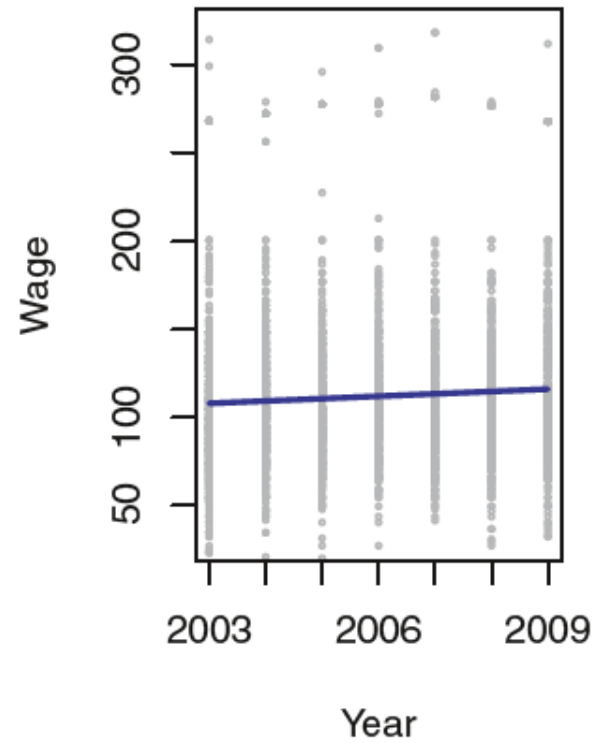
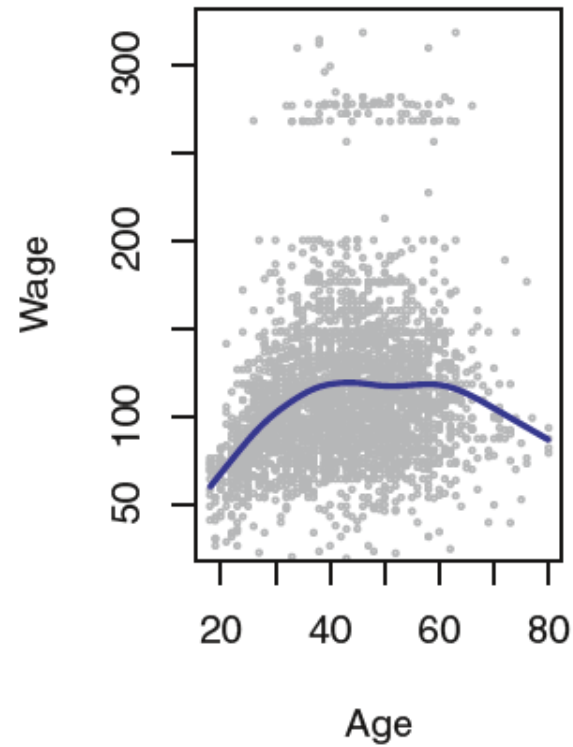
Навчання з учителем: побудова моделі для передбачення чи оцінки результату на основі вхідних даних.

Навчання без учителя: наявними є вхідні дані, але відсутні відповідні результати.

Розглянемо декілька випадків: заробітна плата; фондовий ринок; експресія генів.

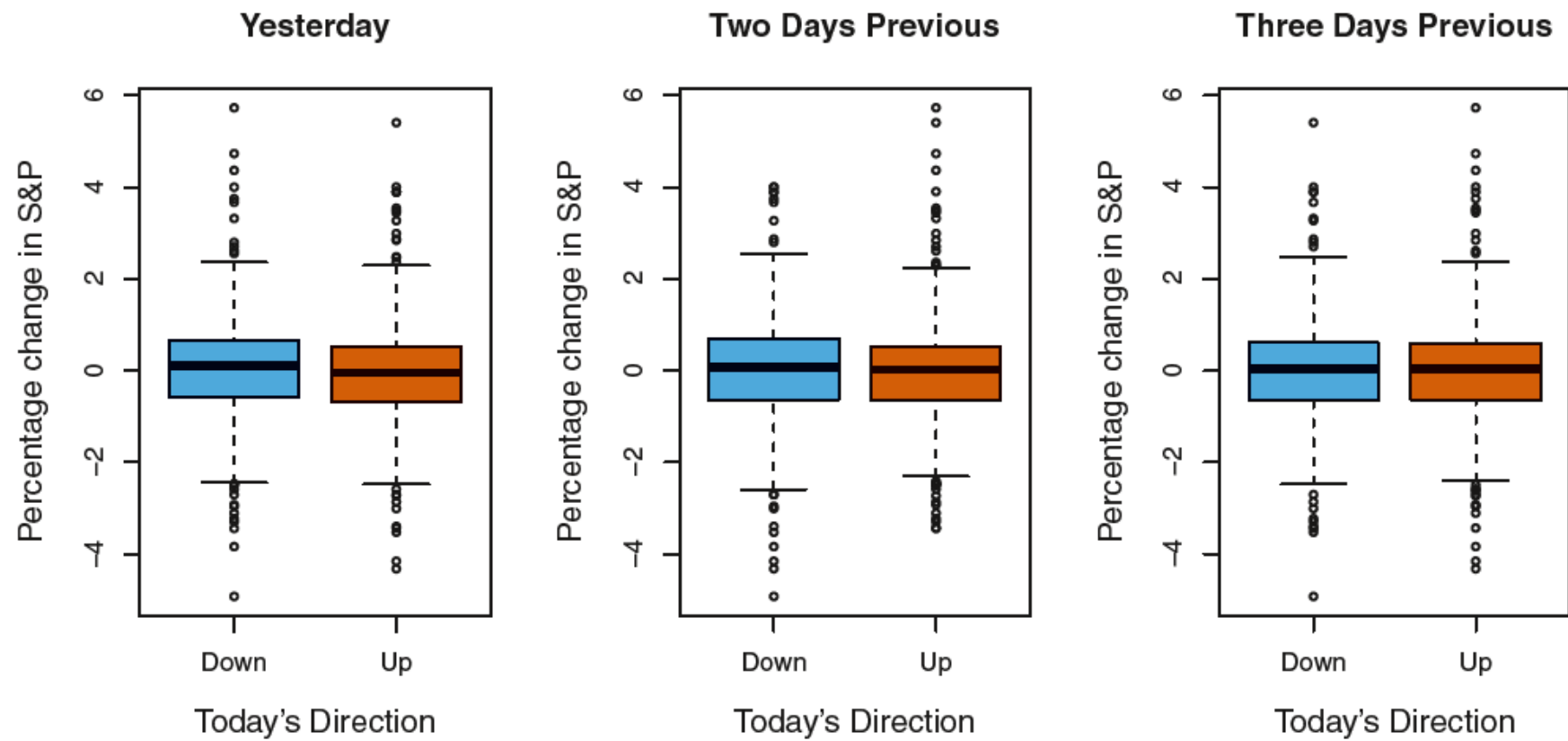
Заробітна плата – дані, що стосуються заробітної плати групи (вибірки) чоловіків з Атлантичного регіону США.

Заробітна плата – дані, що стосуються заробітної плати групи (вибірки) чоловіків з Атлантичного регіону США.



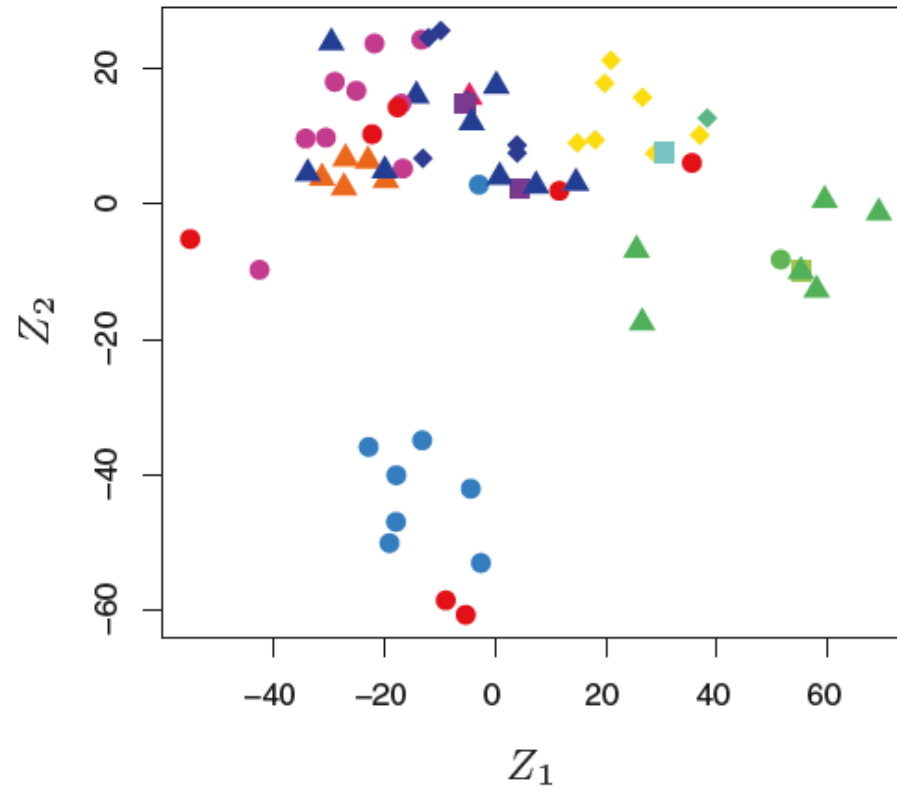
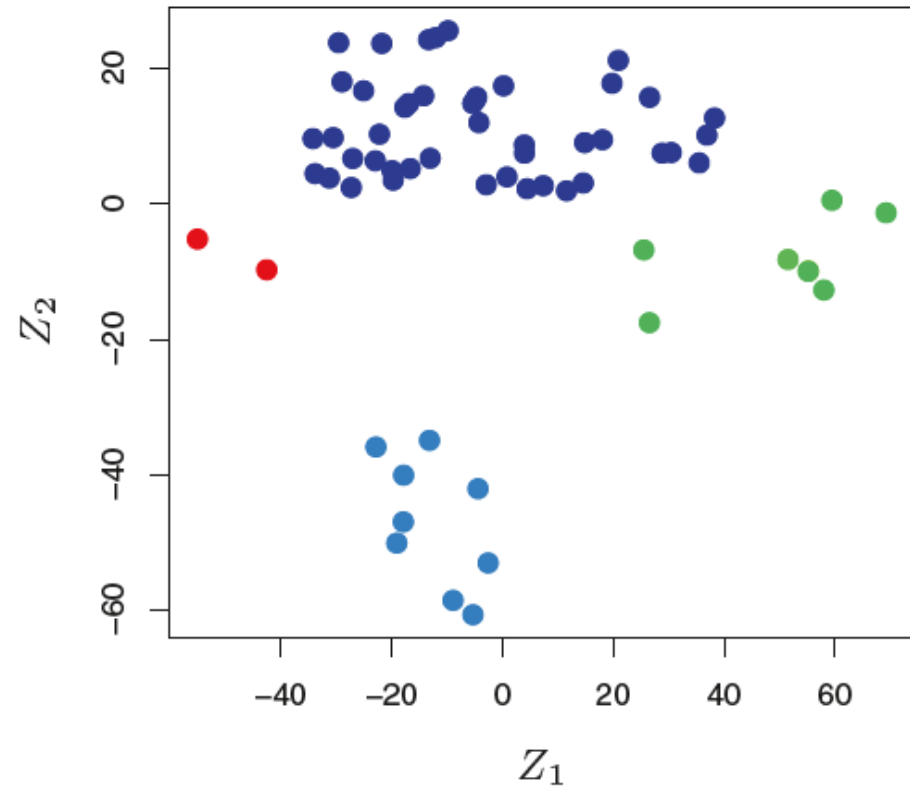
Фондовий ринок – дані, що стосуються щоденної зміни фондового індексу Standard & Poor's 500 (S&P) за 5-річний період між 2001 і 2005 рр.

Фондовий ринок – дані, що стосуються щоденної зміни фондового індексу Standard & Poor's 500 (S&P) за 5-річний період між 2001 і 2005 рр.



Експресія генів – складається з 6830 вимірювань експресії генів для кожної з 64 ракових клітинних ліній.

Експресія генів – складається з 6830 вимірювань експресії генів для кожної з 64 ракових клітинних ліній.



На початку дев'ятнадцятого століття Лежандр і Гаус (Legendre and Gauss) - метод найменших квадратів.

На початку дев'ятнадцятого століття Лежандр і Гаус (Legendre and Gauss) - метод найменших квадратів.

В 1936 році Фішер (Fisher) - лінійний дискримінантний аналіз для прогнозування якісних значень.

На початку дев'ятнадцятого століття Лежандр і Гаус (Legendre and Gauss) - метод найменших квадратів.

В 1936 році Фішер (Fisher) - лінійний дискримінантний аналіз для прогнозування якісних значень.

1940-ті рр.різні автори висували альтернативний підхід - логістичну регресію.

На початку дев'ятнадцятого століття Лежандр і Гаус (Legendre and Gauss) - метод найменших квадратів.

В 1936 році Фішер (Fisher) - лінійний дискримінантний аналіз для прогнозування якісних значень.

1940-ті рр. різні автори висували альтернативний підхід - логістичну регресію.

На початку 1970-х Нелдер та Веддерберн (Nelder and Wedderburn) - узагальнені лінійні моделі.

На початку дев'ятнадцятого століття Лежандр і Гаус (Legendre and Gauss) - метод найменших квадратів.

В 1936 році Фішер (Fisher) - лінійний дискримінантний аналіз для прогнозування якісних значень.

1940-ті рр. різні автори висували альтернативний підхід - логістичну регресію.

На початку 1970-х Нелдер та Веддерберн (Nelder and Wedderburn) - узагальнені лінійні моделі.

В середині 1980-х Брейман, Фрідман, Ольшен і Стоун (Breiman, Friedman, Olshen and Stone) ввели класифікаційні і регресійні дерева.

На початку дев'ятнадцятого століття Лежандр і Гаус (Legendre and Gauss) - метод найменших квадратів.

В 1936 році Фішер (Fisher) - лінійний дискримінантний аналіз для прогнозування якісних значень.

1940-ті рр. різні автори висували альтернативний підхід - логістичну регресію.

На початку 1970-х Нелдер та Веддерберн (Nelder and Wedderburn) - узагальнені лінійні моделі.

В середині 1980-х Брейман, Фрідман, Ольшен і Стоун (Breiman, Friedman, Olshen and Stone) ввели класифікаційні і регресійні дерева.

В 1986 році Хасті і Тібшірані (Hastie and Tibshirani) ввели термін узагальнена адитивна модель для класу нелінійних розширень узагальнених лінійних моделей.

Name	Description
Auto	Gas mileage, horsepower, and other information for cars.
Boston	Housing values and other information about Boston suburbs.
Caravan	Information about individuals offered caravan insurance.
Carseats	Information about car seat sales in 400 stores.
College	Demographic characteristics, tuition, and more for USA colleges.
Default	Customer default records for a credit card company.
Hitters	Records and salaries for baseball players.
Khan	Gene expression measurements for four cancer types.
NCI60	Gene expression measurements for 64 cancer cell lines.
OJ	Sales information for Citrus Hill and Minute Maid orange juice.
Portfolio	Past values of financial assets, for use in portfolio allocation.
Smarket	Daily percentage returns for S&P 500 over a 5-year period.
USArrests	Crime statistics per 100,000 residents in 50 states of USA.
Wage	Income survey data for males in central Atlantic region of USA.
Weekly	1,089 weekly stock market returns for 21 years.