

# Прилагане на обучение чрез минимизация на функция при диагностиране на рак на гърдата

---

Prof. Ludmil Dakovski  
Zekie Shevked

# Обучение от примери

---

## Дадено

- Класифицирани трениращи примери

## Търси се

- Понятие, което представя класа по-компактно

## Цел

- Определяне класа на несрещани примери

# Сложност на задачата

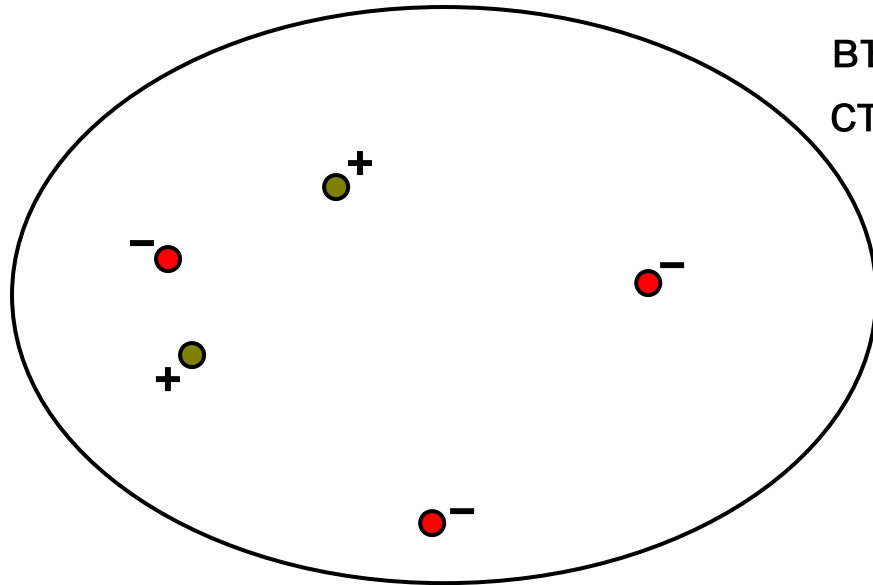
---

- $n$  – брой атрибути
- $k$  – брой стойности за атрибут
- Възможните примери са  $k^n$
- Възможните хипотези са  $2^{k^n}$

При  $n = 5$  и  $k = 2$  се получават  
 $2^{32}$  възможни хипотези!!!

# Подход за правене на изводи

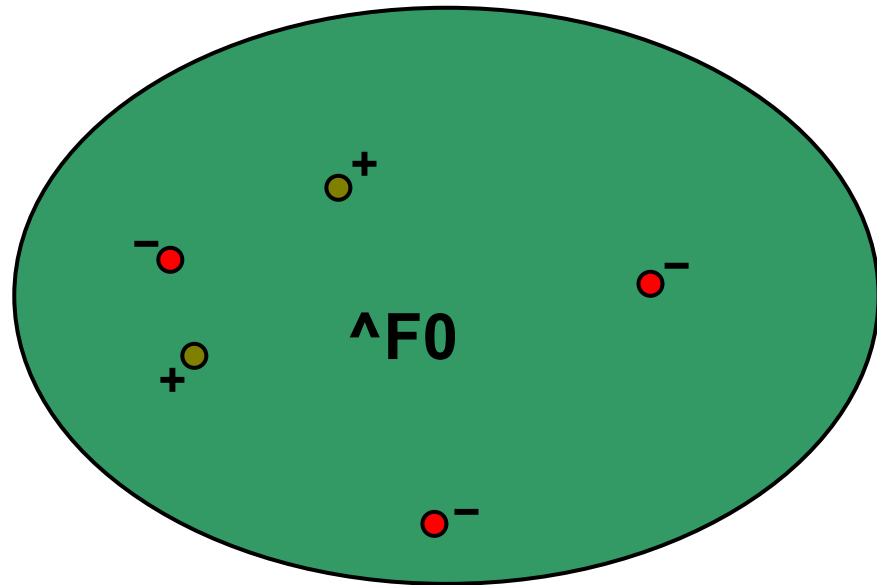
Набор  
ВЪЗМОЖНИ  
СТОЙНОСТИ



$$F1 = P_1 \vee P_2$$

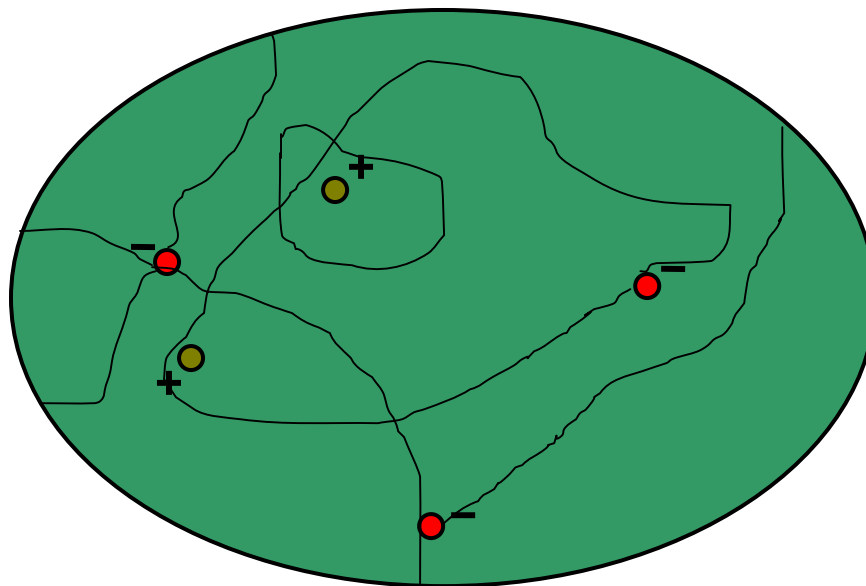
$$F0 = N_1 \vee N_2 \vee N_3$$

# Подход за правене на изводи



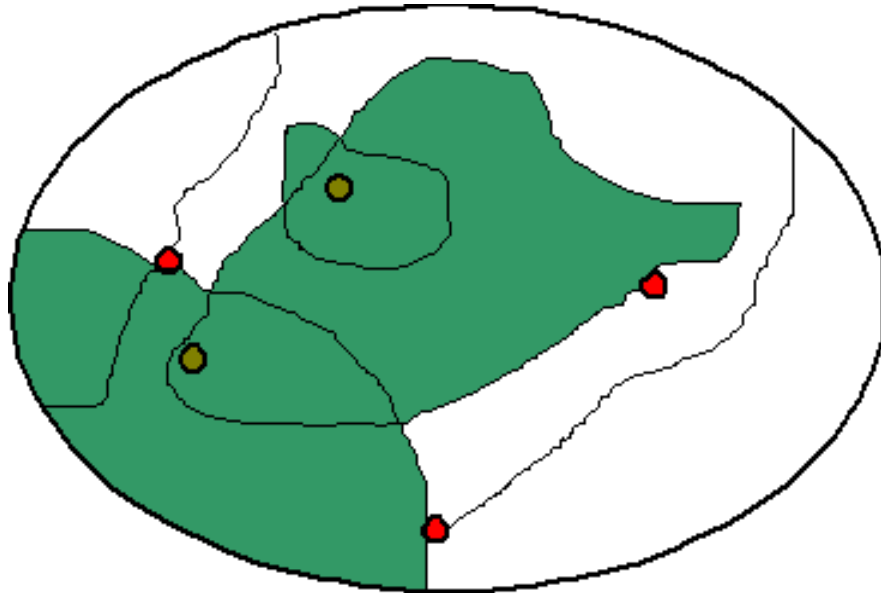
$\hat{F}_0$  покрива всички положителни примери

# Подход за правене на изводи



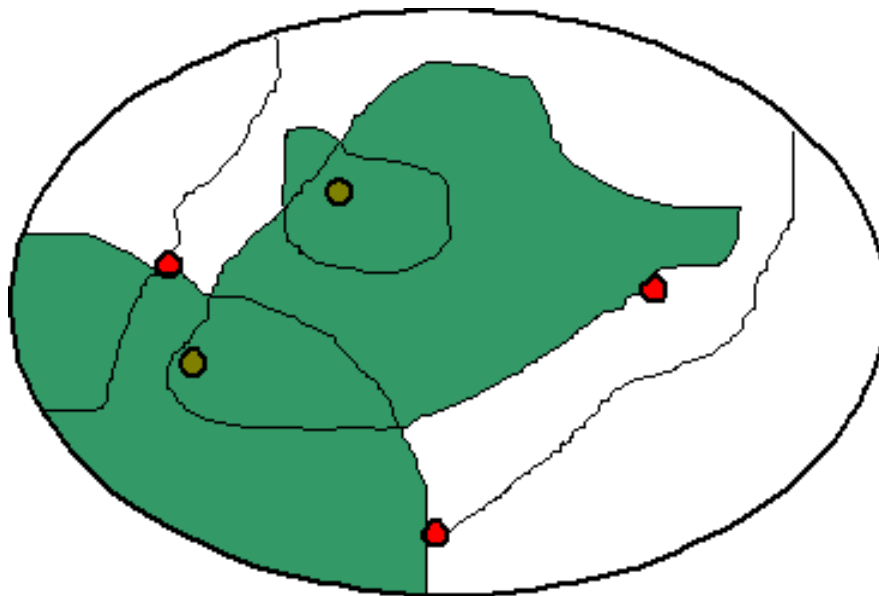
$\hat{F}_0$  се представя по-компактно чрез простите си импликанти

# Подход за правене на изводи



Но само тези импликанти, които покриват положителни примери, са необходими

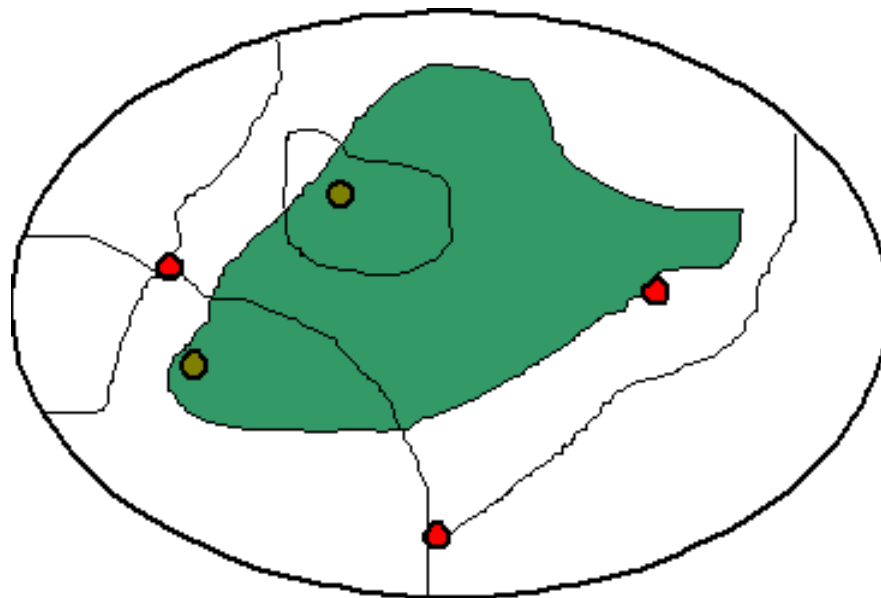
# Настройване получената апроксимация на класа



Настройващ фактор = 1



# Настройване получената апроксимация на класа



Настройващ фактор = 2

# Прилагане на метода при диагностика рак на гърдата

---

- Източник на експерименталните данни

UCI Data Repository

Wisconsin Diagnostic Breast Cancer Data set

Два класа: доброкачествен или злокачествен тумор

Общ брой примери: 569

- 357 доброкачествени
- 212 злокачествени

# Прилагане на метода при диагностика рак на гърдата

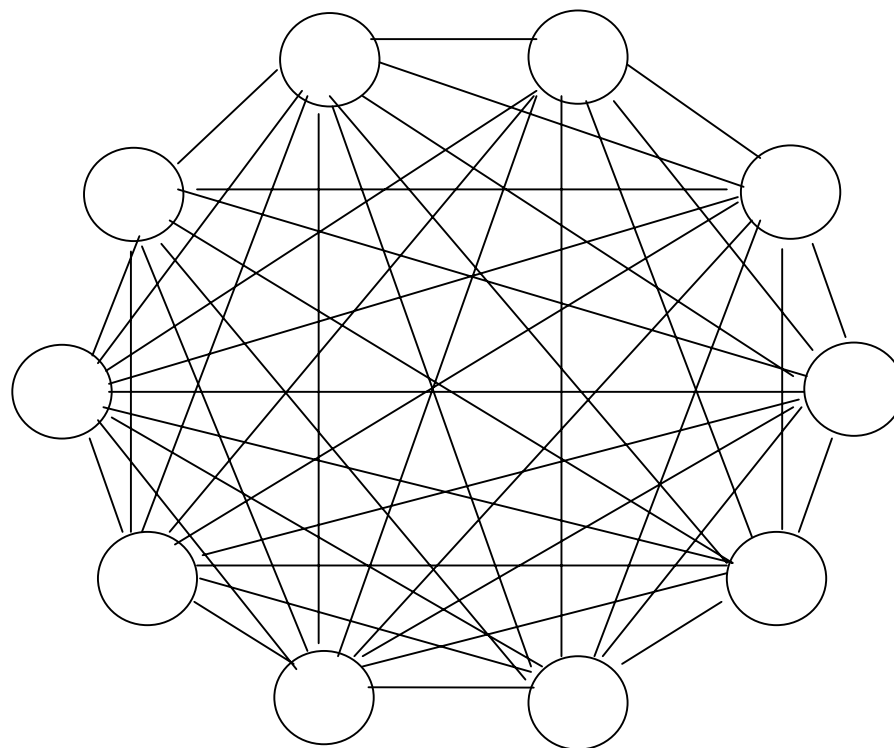
32 на брой атрибута (ID, диагноза и 30 характеристики)  
10 характеристики на ядрото на клетката

- Радиус
- Текстура
- Периметър
- Площ
- Гладкост
- Компактност
- Вдлъбнатост
- Точки на вдлъбнатост
- Симетричност
- Фрактална дименсия

Средна, стандартна и максимална грешка

# 10 fold Cross-validation

---



# Резултати от валидирането

№ на трениращия набор	Най-добър резултат (%)	Настройващ фактор
1	98.25	35
2	96.49	13-17
3	96.49	21-29
4	96.49	12-24
5	96.49	15-24
6	96.49	24-25
7	94.74	26-33
8	94.74	18-24
9	91.23	20-31
10	96.49	16-24

# Заклучение

---

- Нов алгоритъм за минимизация
- Позволява допълнителна настройка
- Висок процент на точност при класифициране
- Подходящ за диагностика