Announcements

- Final exam
  - Aug. 6, 8am-9:40am
  - Open book, open notes
  - No communication
- Still some questions?
  - Piazza
  - OH Mon 9am-10am, Wed 9am-10am
- Course evaluation

Ve492: Introduction to Artificial Intelligence
Final Review

Paul Weng

UM-SJTU Joint Institute

Slides adapted from http://ai.berkeley.edu, AIMA, UM, CMU

Probability

- For each of the following statements, either prove it is true or give a counterexample.

1) If P(a | b, c) = P(b | a, c), then P(a | c) = P(b | c) True

2) If P(a | b, c) = P(a), then P(b | c) = P(b) False

3) If P(a | b) = P(a), then P(a | b, c) = P(a | c) False

1) $P(a \mid \cancel{b}, c) = \dfrac{P(\cancel{b \mid} a, c) \cdot P(a \mid c)}{P(b \mid c)}$ by Bayes rule

2) $a \perp\!\!\!\perp \{b, c\} \Rightarrow b \perp\!\!\!\perp c$? False

3) $a \perp\!\!\!\perp b \Rightarrow a \perp\!\!\!\perp b \mid c$?

$$\begin{tikzcd} a \arrow[dr] & & b \arrow[dl] \\ & c & \end{tikzcd}$$

False

Bayes Rule

$P(A|V)$

Consider two medical tests, A and B, for a virus. Test A is 95% effective at recognizing the virus when it is present, but has a 10% false positive rate (indicating that the virus is present, when it is not). Test B is 90% effective at recognizing the virus, but has a 5% false positive rate. The two tests use independent methods of identifying the virus. The virus is carried by 1% of all people. Say that a person is tested for the virus using only one of the tests, and that test comes back positive for carrying the virus.

$P(A|\bar{V})$ $P(B|V)$ $P(B|\bar{V})$ $P(V)$

- Which test returning positive is more indicative of someone really carrying the virus? Justify your answer mathematically.

A : test A is positive B : test B is positive

V : patient has virus $P(V|A) \overset{?}{} P(V|B)$

$$P(V|A) = \frac{P(A|V) \cdot P(V)}{P(A)} = \frac{P(A|V) \cdot P(V)}{P(A|V) \cdot P(V) + P(A|\bar{V}) P(\bar{V})} \approx 8.8\%$$

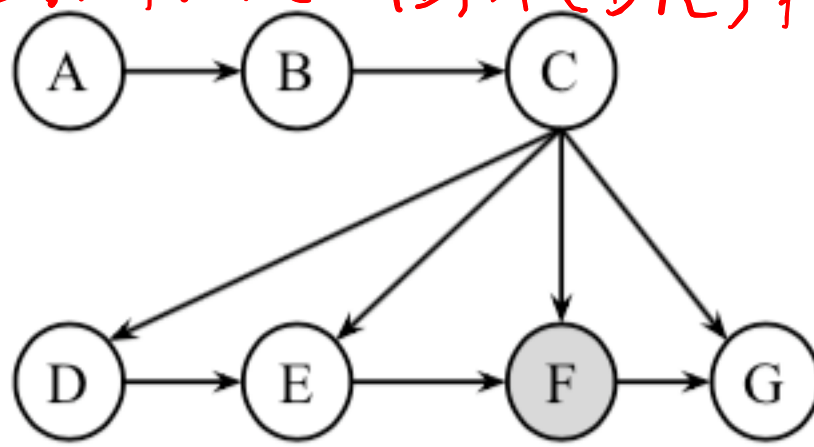
$P(V|B) \approx 15.4\%$

Bayes' Net

domain size $= d$.

- Write the joint distribution of the following Bayes' net

$P(A, B, \dots, G) = P(A) \cdot P(B|A) \cdot P(C|B) \cdot P(D|C)\, P(E|C,D)\, P(F|C,E)\, P(G|C,F)$



- How many values does the joint distribution have? $d^7 - 1$

- How many parameters does the Bayes' net have?

$d + d^2 + d^2 + d^2 + d^3 + d^3 + d^3$

$\Rightarrow (d-1) + 3d \times (d-1) + 3d^2 (d-1)$

D-Separation

- Check Bayes applet

Inference

A] $P(A)\,P(B|A) \xrightarrow{\text{join}} f_0(A,B)$

$\xrightarrow{\text{eliminate}} f_1(B)$

C] $P(C|B)\,P(D|C)\,P(E|C,D)\,P(+f|C,E)\,P(G|+f,C)$

$\rightarrow f_2(B,C,D,E,+f,G)$

$\rightarrow f_3(B,D,E,+f,G)$

E] $f_3(B,D,E,+f,G)$

$\rightarrow f_4(B,D,+f,G)$

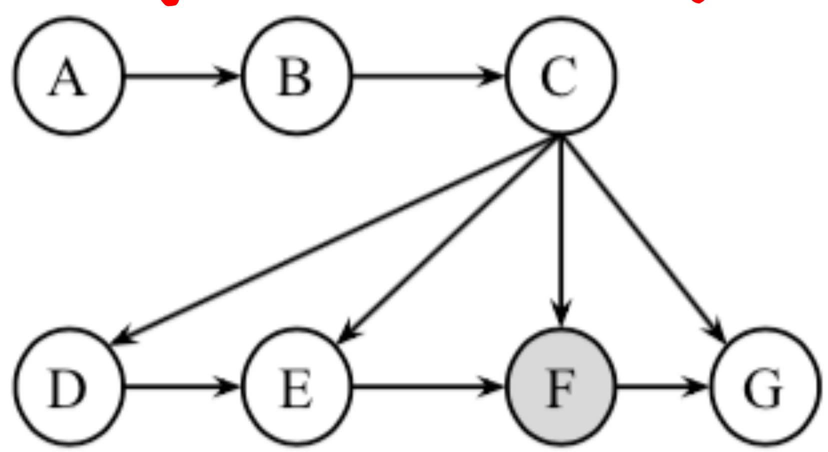
G] $f_4(B,D,+f,G) \rightarrow f_5(B,D,+f)$

$f_1(B)\,f_5(B,D,+f)$

$\rightarrow f_6(B,D,+f)$

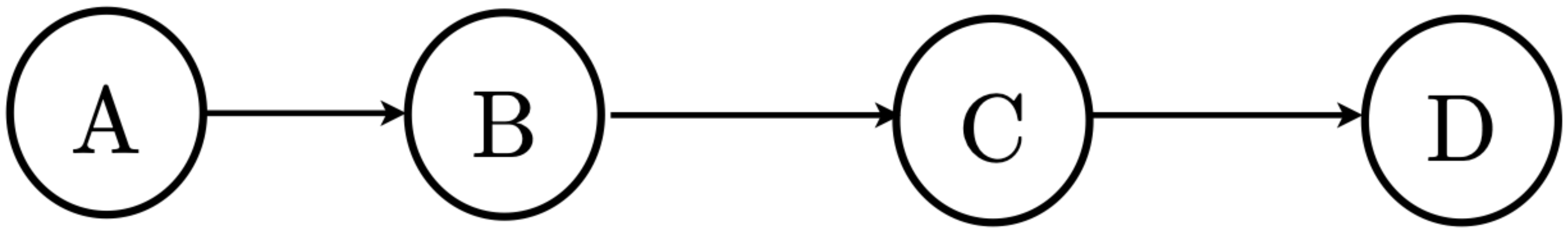
$\Downarrow$ normalize

$P(B,D|+f)$



- Run Variable Elimination to compute $P(B, D| + f)$ with order A, C, E, G ✓
- What is the size of the largest generated factor? 5
- Find the best ordering for Variable Elimination G, E, C, A
  A, G, E, C
- What is the cutset for this graph? {C}

Sampling



$P(A)$	
$-a$	$3/4$
$+a$	$1/4$

$P(B\|A)$		
$-a$	$-b$	$2/3$
$-a$	$+b$	$1/3$
$+a$	$-b$	$4/5$
$+a$	$+b$	$1/5$

$P(C\|B)$		
$-b$	$-c$	$1/4$
$-b$	$+c$	$3/4$
$+b$	$-c$	$1/2$
$+b$	$+c$	$1/2$

$P(D\|C)$		
$-c$	$-d$	$1/8$
$-c$	$+d$	$7/8$
$+c$	$-d$	$5/6$
$+c$	$+d$	$1/6$

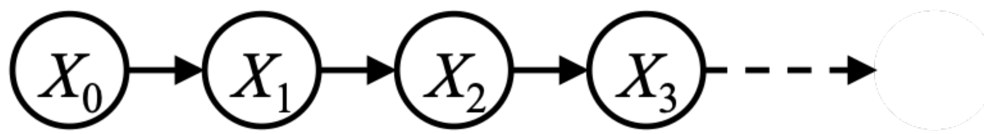
Samples

$\frac{1}{5}\cdot\frac{1}{8}$	$+a$	$+b$	$-c$	$-d$
$\frac{1}{5}\cdot\frac{5}{6}$	$+a$	$-b$	$+c$	$-d$
$\frac{1}{3}\cdot\frac{5}{6}$	~~$-a$~~	~~$+b$~~	~~$+c$~~	~~$-d$~~
$\frac{1}{3}\cdot\frac{5}{6}$	~~$-a$~~	~~$-b$~~	~~$+c$~~	~~$-d$~~
	~~$+a$~~	~~$-b$~~	~~$-c$~~	~~$+d$~~
	$+a$	$+b$	$+c$	$-d$
	~~$-a$~~	~~$+b$~~	~~$-c$~~	~~$+d$~~
	~~$-a$~~	~~$-b$~~	~~$+c$~~	~~$-d$~~

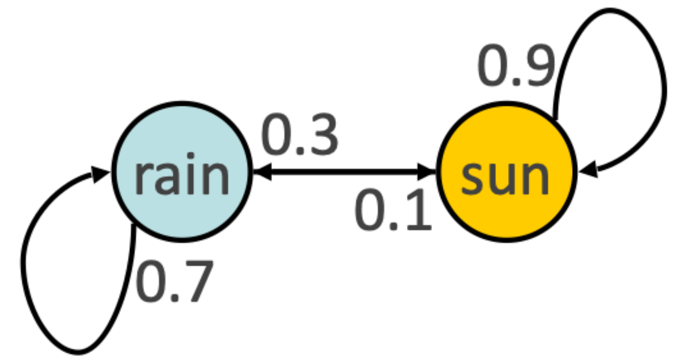
- Estimate $P(+c|+a,-d)$ via rejection sampling $\frac{2}{3}$
- Estimate $P(-a|+b,-d)$ via likelihood weighting

$$\frac{\left(\frac{1}{3}\cdot\frac{5}{6}+\frac{1}{3}\cdot\frac{5}{6}\right)}{\frac{1}{3}\cdot\frac{5}{6}+\frac{1}{3}\cdot\frac{5}{6}+\frac{1}{5}\cdot\frac{1}{8}+\frac{1}{5}\cdot\frac{5}{6}}$$

Markov Chain



$P(X_0)$ $P(X_t \mid X_{t-1})$

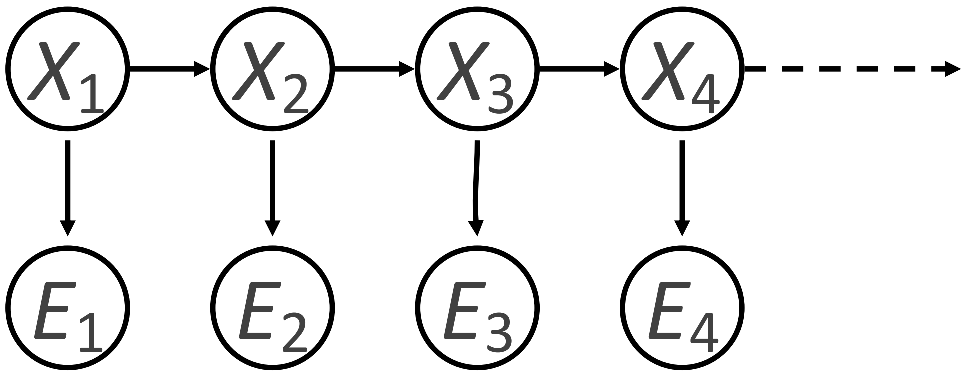


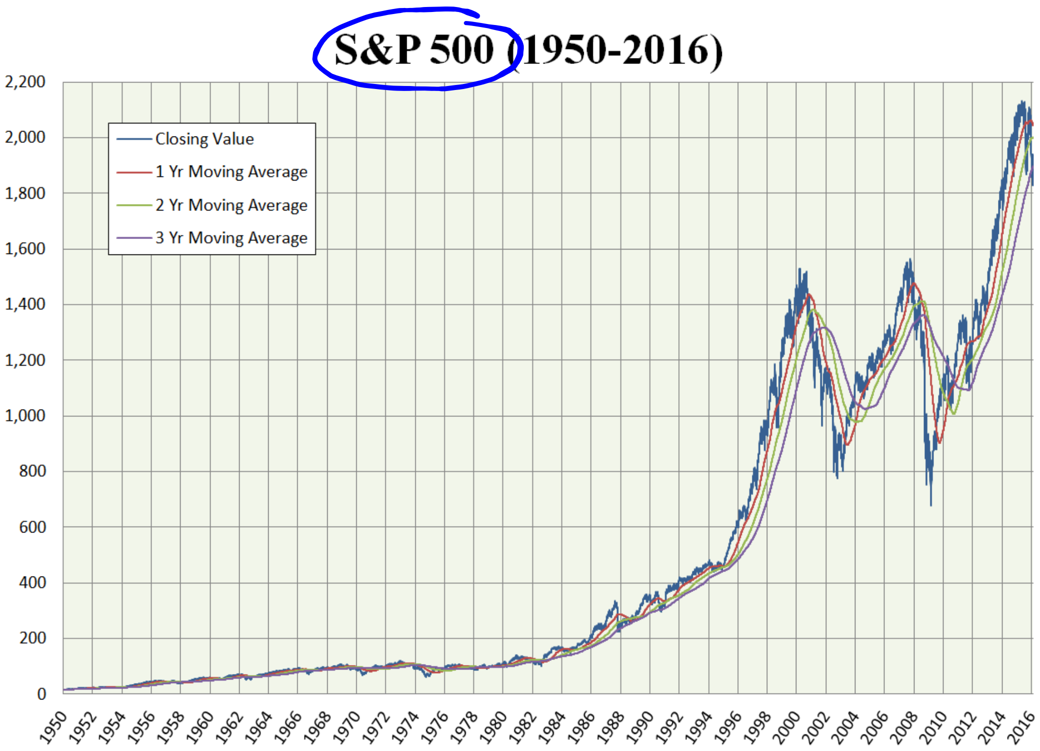
$P(X_t) = \sum_x P(X_t \mid X_{t-1}=x) P(X_{t-1}=x)$

- What is the probability of $P(X_t)$? $P(X_0)$
- What is the stationary distribution for the weather example?

$$P_\infty = \begin{array}{c c} & \begin{array}{cc} \text{rain} & \text{sun} \end{array} \\ \begin{array}{c} \text{rain} \\ \text{sun} \end{array} & \begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{pmatrix}^T \end{array} P_\infty$$

Hidden Markov Model

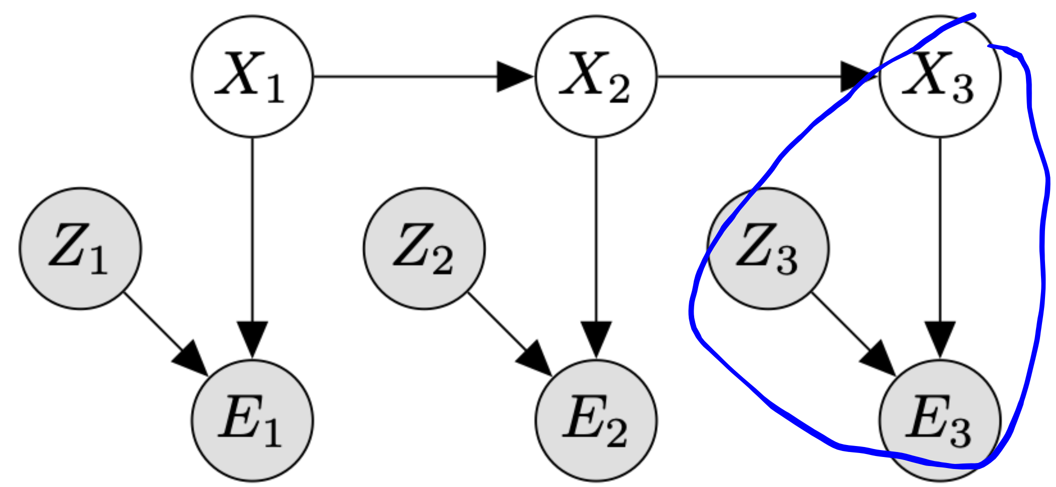




Wikipedia

- Financial investment
  - X = market condition: bull, bear
  - E = price evolution of some index: up, down
- Label past data into bull vs bear
- Use historical data to estimate transition/emission probabilities

Hidden Markov Model

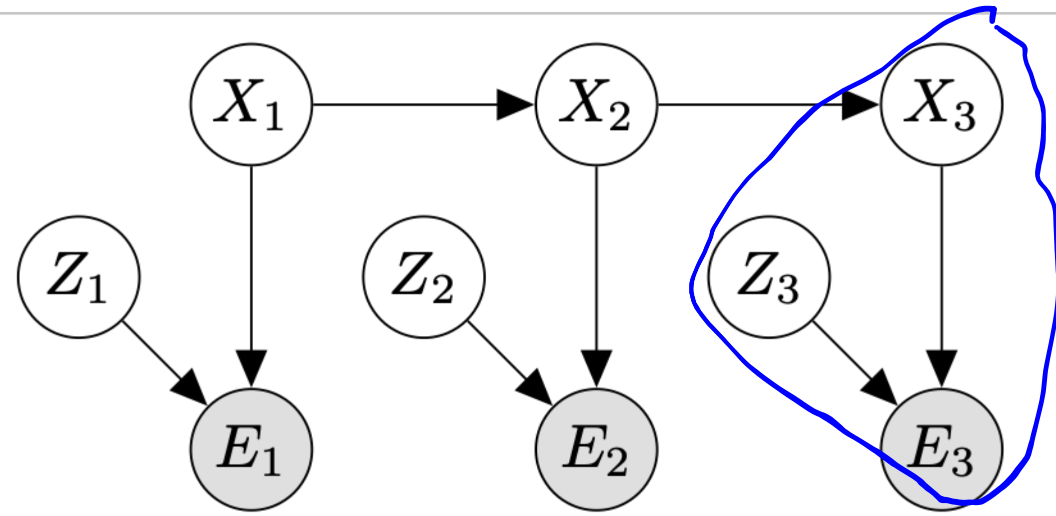


- Adapt the forward algorithm to this variant of HMM
  - Predict step $P(X_{t+1} \mid \underline{z_{1:t}}, e_{1:t}) = \sum_x P(X_{t+1} \mid X_t = x) \cdot \overbrace{P(X_t = x \mid z_{1:t}, e_{1:t})}^{\text{belief}}$
  - Update

$$P(X_{t+1} \mid z_{1:t+1}, e_{1:t+1}) \propto P(Z_{t+1} = z_{t+1}) \cdot P(X_{t+1} \mid z_{1:t}, e_{1:t}) \cdot$$

$$P(E_{t+1} = e_{t+1} \mid z_{t+1}, X_{t+1})$$

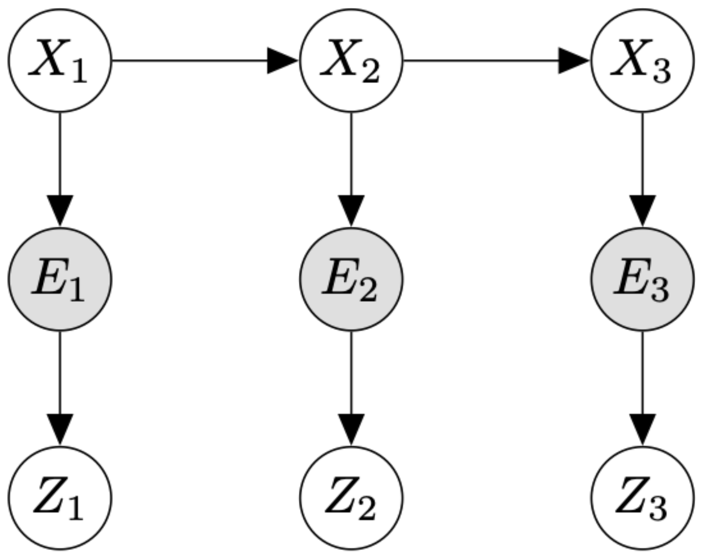
Hidden Markov Model



- Adapt the forward algorithm to this variant of HMM
  - Predict step $P(X_{t+1} \mid e_{1:t}) = \sum_{x_t} P(X_{t+1} \mid x_t) \cdot \overbrace{P(x_t \mid e_{1:t})}^{\text{belief}}$
  - Update

$$P(X_{t+1} \mid e_{1:t+1}) \propto P(X_{t+1} \mid e_{1:t}) \cdot \sum_{z_{t+1}} P(e_{t+1} \mid X_{t+1}, z_{t+1}) P(z_{t+1})$$

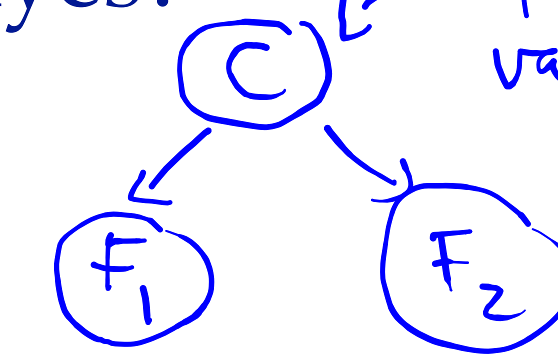
Hidden Markov Model



- Adapt the forward algorithm to this variant of HMM
  - Predict step
  - Update

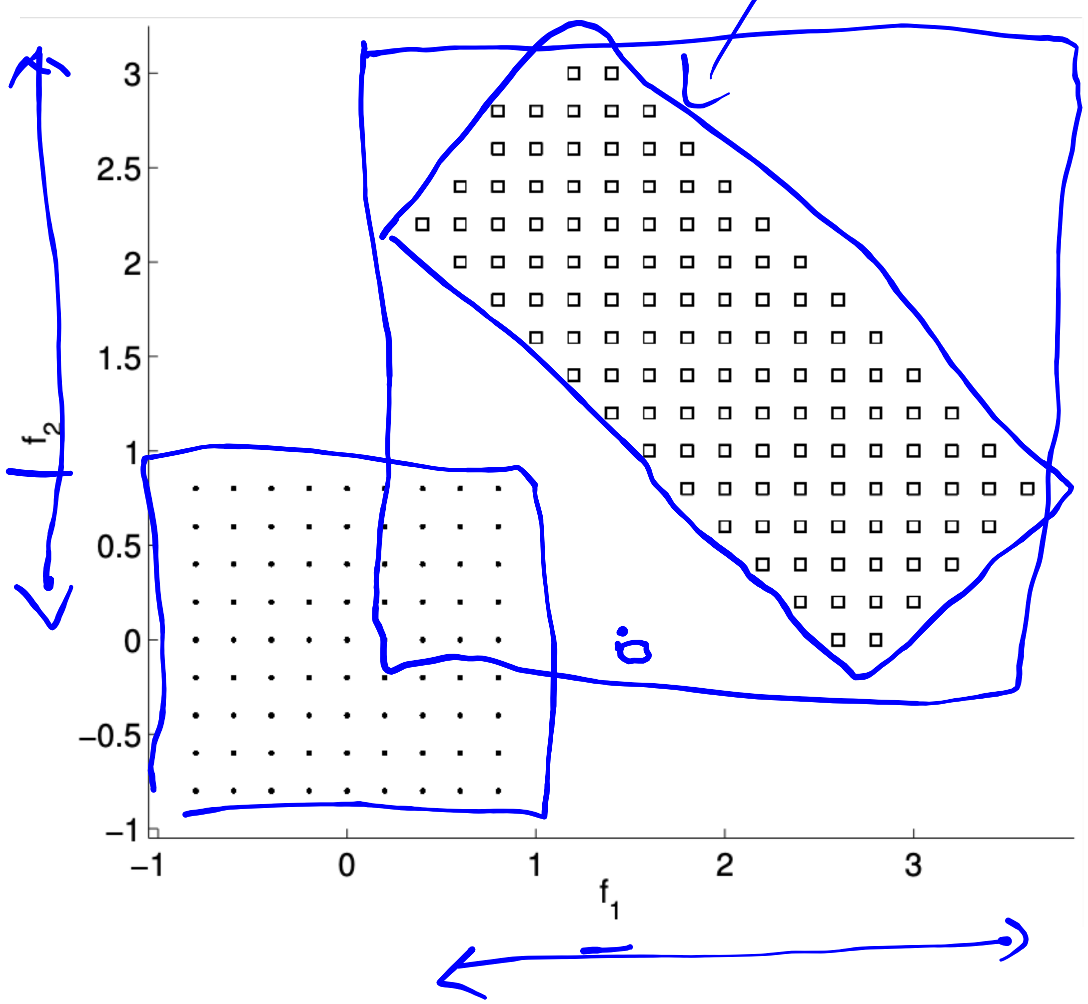
Naïve Bayes

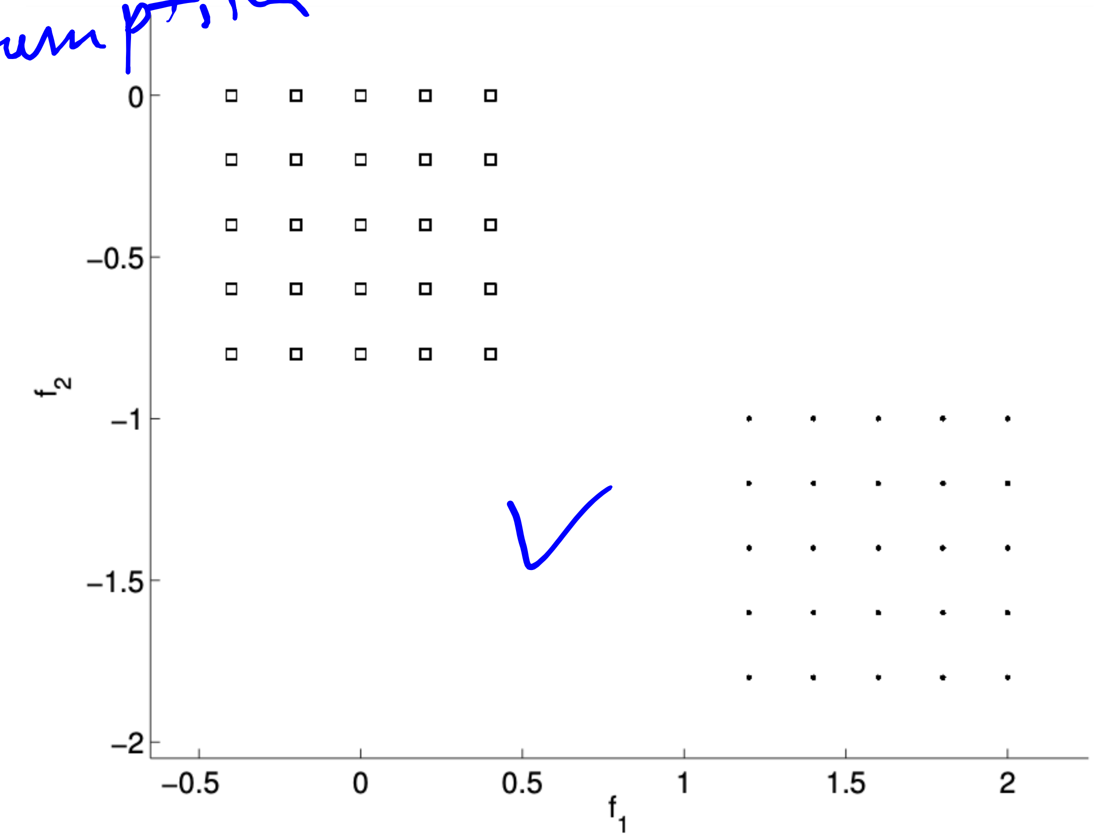
- Which of the following binary classification problems satisfy the assumption made in Naïve Bayes?



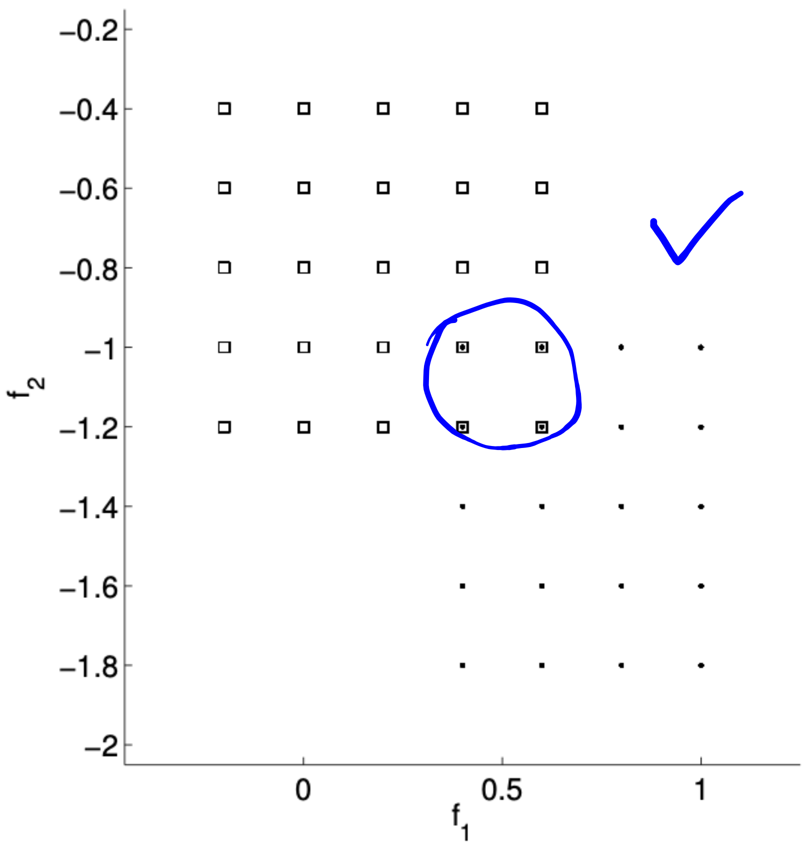
2 possible values.

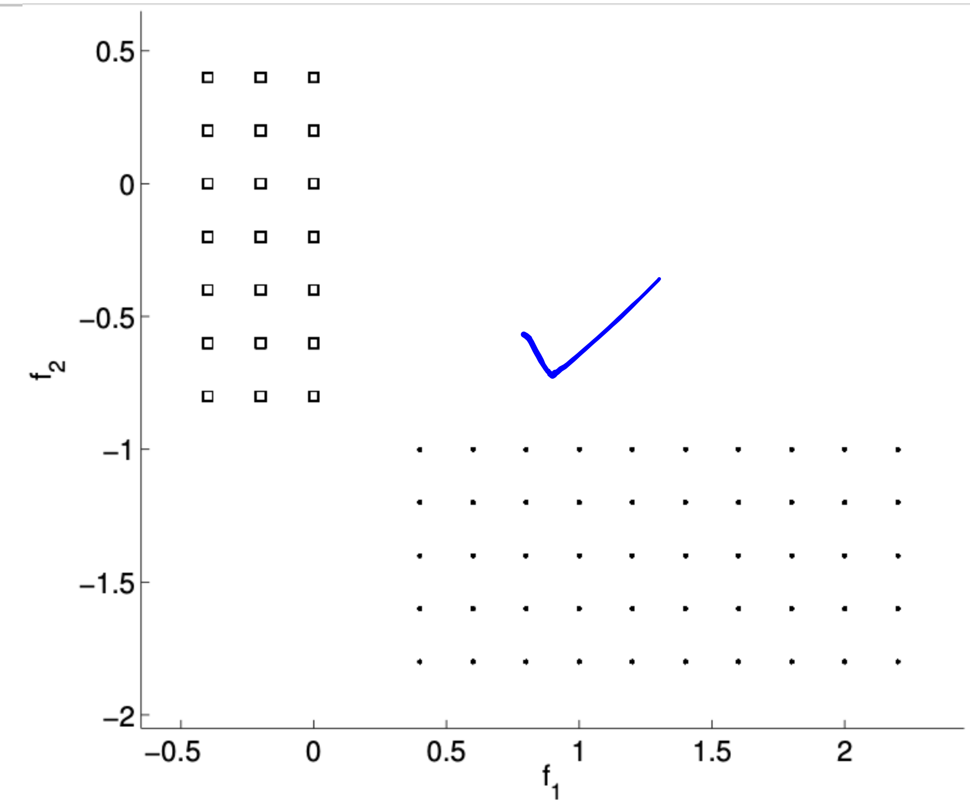
class violates Naïve Bayes assumption

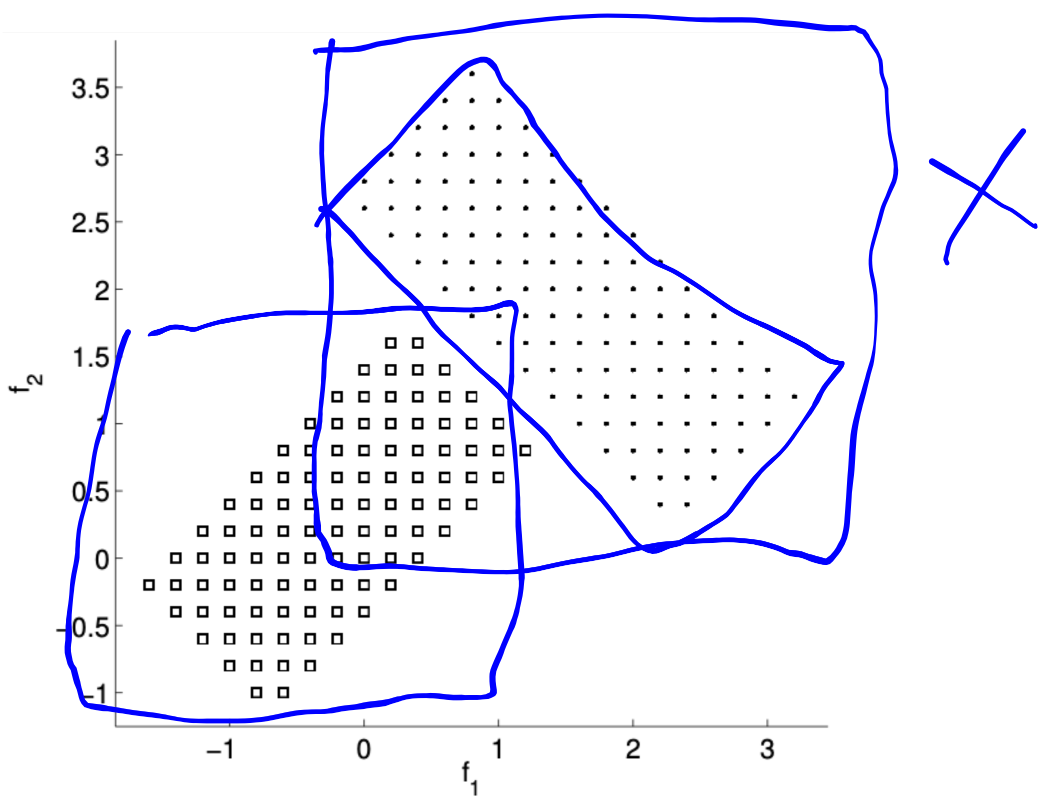


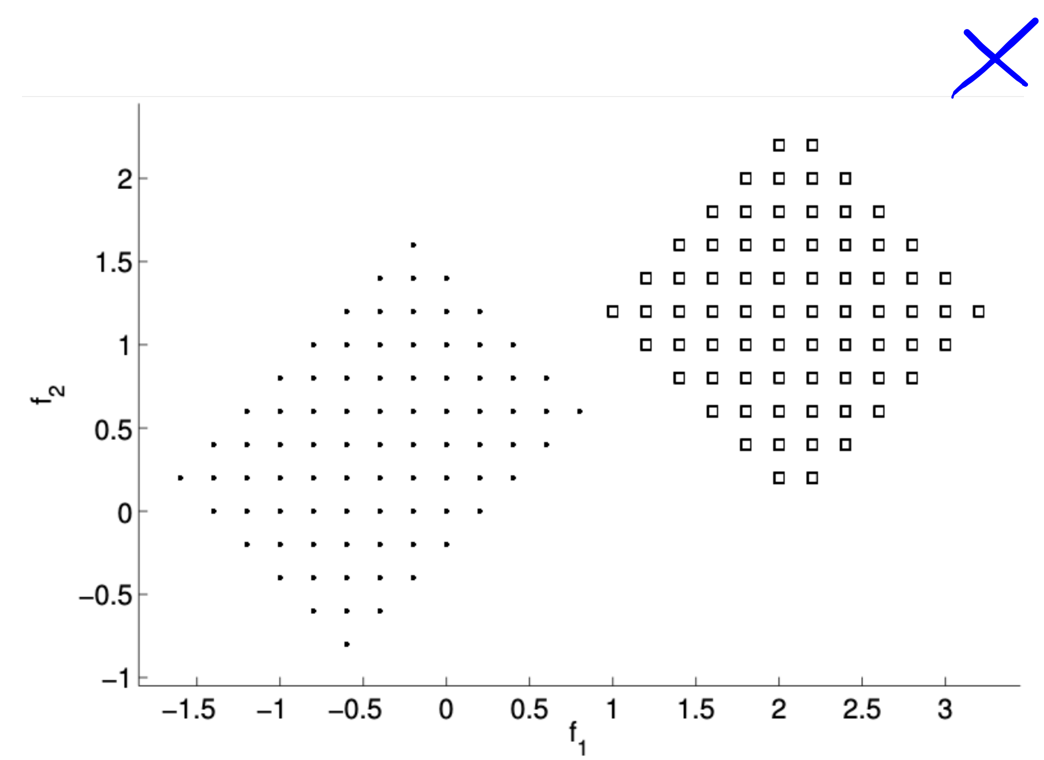


Naïve Bayes ctd.









Discriminative Learning

For a binary classification problem, we choose the following model $\mathbb{P}(y = +1|x) = \Phi(w \cdot x)$ where Φ is the CDF of a standard normal distribution.

$P(X \le x)$

1) What is the decision boundary?

2) Formulate the optimization problem to be solved to find w

3) Formulate the stochastic gradient method to solve that problem

1) $P\{y=+1|x\} = P(y=0|x)$

$\Phi(w \cdot x) = 1 - \Phi(w \cdot x)$

$\Phi(w \cdot x) = 1/2$

$w x = \Phi^{-1}(1/2) = 0$

$w x = 0$

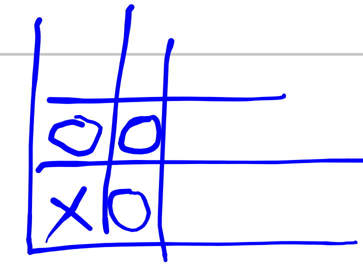
2) $D = \{(x_i, y_i) \mid i = 1 \ldots N\}$

Write likelihood of D assuming iid

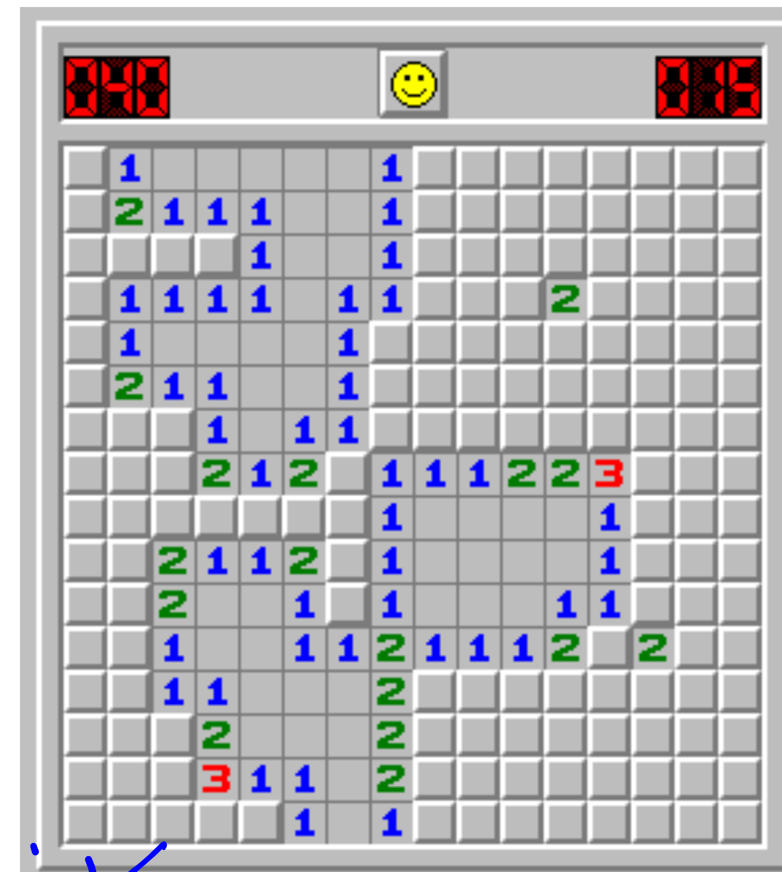
Propositional Logic

- Which of the following are correct? $\alpha \vDash B$ iff $model(\alpha) \subseteq model(B)$
  - ✓ False ⊨ True $model(False) = \phi$
  - ✗ True ⊨ False
  - ✓ $(A \wedge B) \vDash (A \Leftrightarrow B)$ $model(A \wedge B) = \{(A = T, B = T)\} \rightarrow size = 1$
  - $(A \vee B) \wedge (\neg C \vee \neg D \vee E) \vDash (A \vee B)$
  - $(A \vee B) \wedge \neg(A \Rightarrow B)$ is satisfiable
  - $(A \Leftrightarrow B) \wedge (\neg A \vee B)$ is satisfiable
  - $(A \Leftrightarrow B) \Leftrightarrow C$ has the same number of models as $(A \Leftrightarrow B)$ for any fixed set of proposition symbols that includes A, B, C

Application: Propositional Logic



- Minesweeper: Let $X_{i,j}$ be true iff square $[i, j]$ contains a mine.
  - Write down the assertion that exactly two mines are adjacent to [1,1] as a sentence involving some logical combination of $X_{i,j}$ propositions. $X_{1,2}$ $X_{2,1}$ $X_{2,2}$
  - Generalize your assertion by explaining how to construct a CNF sentence asserting that k of n neighbors contain mines
  - Explain precisely how an agent can use DPLL to prove that a given square does (or does not) contain a mine.



$KB \models X_{i,j}$ ✓

$KB \wedge \neg X_{i,j} \rightarrow CNF \rightarrow DPLL$

$KB \models \neg X_{i,j}$ ✓

$(X_{1,2} \wedge X_{2,1} \wedge \neg X_{2,2}) \vee (X_{1,2} \wedge \neg X_{2,1} \wedge X_{2,2}) \vee (\neg X_{1,2} \wedge X_{2,1} \wedge X_{2,2})$

First-Order Logic

- For each of the following sentences in English, decide if the accompanying first-order logic sentence is a good translation. If not, explain why not and correct it.

  - No two people have the same social security number.

    $\neg\exists$ x, y, n Person(x) $\wedge$ Person(y) $\Rightarrow$ [HasSS#(x, n) $\wedge$ HasSS#(y, n)]

    $\forall x, y$ Person(x) $\wedge$ Person(y) $\Rightarrow \neg\exists n$ HasSS#(x, n) $\wedge$ HasSS#(y, n)

  - John's social security number is the same as Mary's.

    ✓ $\exists$ n HasSS#(John, n) $\wedge$ HasSS#(Mary, n)

  - Everyone's social security number has nine digits.

    $\forall$ x, n Person(x) $\Rightarrow$ [HasSS#(x, n) $\wedge$ Digits(n, 9)]

    $\forall x$ Person(x) $\Rightarrow \exists n$ HasSS#(x, n) $\wedge$ Digits(n, 9)

    $\forall n$ HasSS#(x, n) $\wedge$ Digits(n, 9)

Classic Planning

A monkey is in a room with some bananas hanging out of reach from the ceiling. A box is available that will enable the monkey to reach the bananas if he climbs on it. Initially, the monkey is at *A*, the bananas are at *B*, and the box is at *C*. The monkey and the box have height *Low*, but if the monkey climbs onto the box he will have height *High*, the same as the bananas. The actions available to the monkey include *EatBananas* if the monkey and the bananas are at the same location and height, *Go* from one place to another, *Push* an object from one place to another, and *ClimbUp* onto or *ClimbDown* from an object.

EatBananas(loc) : Precond : Location(loc)
At(M, loc) At(Ba, loc)
Height(M, H)

1) Write down the initial state description

2) Write down the STRIPS definitions of the five actions.

1) At(monkey, A) At(Bananas, B) At(Box, C)
Height(M, L) Height(Ba, H) Height(Bo, L)
Location(A) Location(B) Location(C)

Effects :
DEL : At(Ba, B)
Height(Ba, H)