Ex. 1   6 min

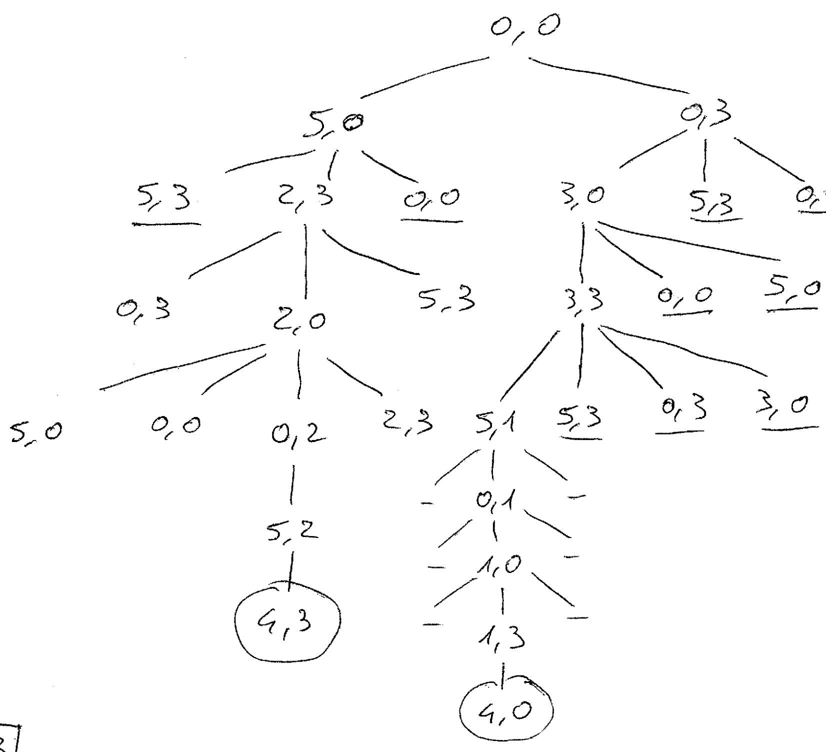
representation: 2 digits for the liters in the 2 containers

init. state: 0,0

operators (successor function):
- fill 5-lit. container
- fill 3-lit cont.
- empty 5-lit cont.
- empty 3-lit cont
- pour 3 in 5
- pour 5 in 3

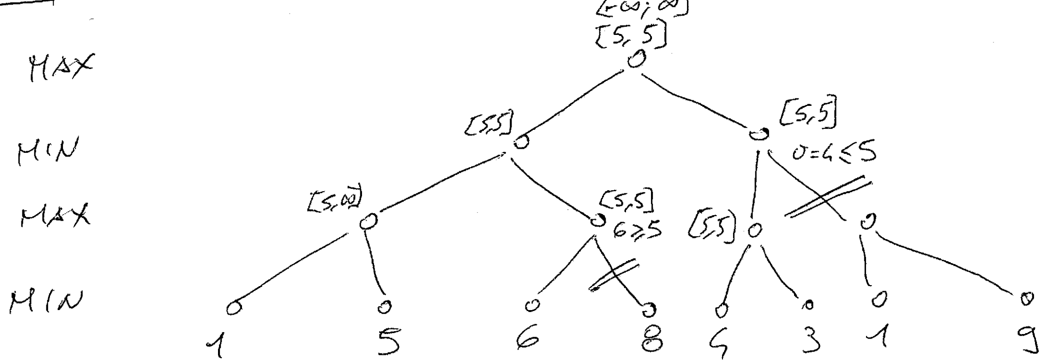
goal state   4,*

Search tree:



— pruned because repeated state.

Ex. 3



Ex. 2 20 min

Unif. cost

A	I (6) B (8)
I	B (8) M (8) J (9)
B	M (8) J (9) F (15) C (18)
M	J (9) N (10) F (15) O (17) C (18)
J	N (10) L (11) F (15) O (17) C (18) K (19)
N	L (11) F (15) O (17) C (18) K (19)
L	F O C K
F	G (17) O (17) C (18) H (18) K (19)
G	O C H K
O	C H K
C	H (18) K (19) E (23) D (28)
H	K D E
K	E D
E	D
D	

Best Cost

A	I (6+6=12) B (8+7=15)
I	J (10) B (15) M (16)
J	B (15) M (16) L (19) K (22)
B	M (16) L (19) K (22) F (25) C (26)
M	N (17) O (18) L (19) K (22) F (25) C (26)
N	O L K F C
O	L K F C
L	K F C
K	F C
F	~~[illegible]~~ H (23) C (26) G (27)
H	C G
C	~~[illegible]~~ E (24) G (27) D (36)
E	G D
G	D
D	

Greedy

A
I
J
K

we cannot say that the best Cost heuristic is an A* search unless we have information on the type of heuristic used. If the heuristic is admissible then we can claim that the best cost heuristic is an A* search

ES.4 — 45 min

(a)
i) $\forall x, y, z \quad Owns(x,y) \land In(z,y) \land File(z) \Rightarrow Owns(x,z) \land File(z)$

ii) $\forall x \quad Ne(x) \Leftrightarrow \exists z \; File(z) \land In(z,x)$

iii) $\exists x \quad Ne(x) \land Owns(J,x)$

iv) $\exists z \quad File(z) \land Owns(J,z)$

b) We can show that $KB \models \alpha$ by showing that $KB \land \neg\alpha$ leads to a contradiction. To prove $KB \land \neg\alpha$ false we use FOL inference rules: transform in CNF, apply resolution with unification, show that the resolution closure contains the empty clause.
If $KB \models \alpha$ holds then the set of interpretations under which KB holds is a subset of the set of interpretations under which α holds.

c) From the last sentence in (b) it follows that we can show $KB \not\models \alpha$ by showing that there exists an interpretation under which α holds but KB not.
Showing that the procedure via resolution of (b) cannot reach a contradiction does not work in general because it may not terminate due to semidecidability of FOL.

d) Transform i-iv in CNF:

i): $\forall x, y, z \; \neg[Owns(x,y) \land In(z,y)] \lor Owns(x,z)$

$\neg Owns(x,y) \lor \neg In(z,y) \lor Owns(x,z)$ (1)

ii): • $\forall x, \; Ne(x) \Rightarrow \exists z \; File(z) \land In(z,x)$

$\forall x \; \neg Ne(x) \lor (File(f(x)) \land In(f(x),x))$

$(\neg Ne(x) \lor File(f(x)))$ (2) $\land$ $(\neg Ne(x) \lor In(f(x),x))$ (3)

• $\forall x \; [\exists z \; File(z) \land In(z,x)] \Rightarrow Ne(x)$

$\forall x \; \neg[\exists z \; File(z) \land In(z,x)] \lor Ne(x)$

$\forall x \; (\forall z \; \neg File(z) \lor \neg In(z,x)) \lor Ne(x)$

$\neg File(z) \lor \neg In(z,x) \lor Ne(x)$ (4)

iii): $Ne(H)$ (5) $\land$ $Owns(J,H)$ (6)

→

¬iv): $\neg \exists z \; File(z) \wedge Owns(J,z)$

$\forall z \; \neg File(z) \vee \neg Owns(J,z)$

$\neg File(z) \vee \neg Owns(J,z)$ (7)

Hence we have 7 clauses.

7: $\neg File(z) \vee \neg Owns(J,z)$

6: $Owns(J,H)$

---

8 ∴ $\neg File(H)$ $\theta: \{z/H\}$

2: $\neg Ne(x) \vee File(f(x))$

5: $Ne(H)$

---

9 ∴ $File(f(H))$ $\theta \{x/H\}$

3: $\neg Ne(x) \vee In(f(x), x)$

5: $Ne(H)$

---

10 ∴ $In(f(H), H)$ $\theta: \{x/H\}$

7: $\neg File(z) \vee \neg Owns(J,z)$

9: $File(f(H))$

11: ∴ $\neg Owns(J, f(H))$ $\theta \{z/f(H)\}$

1: $\neg Owns(x,y) \vee \neg In(z,y) \vee Owns(x,z)$

6: $Owns(J,H)$

---

12 ∴ $\neg In(z,H) \vee Owns(J,z)$ $\theta \{x/J, y/H\}$

12: $\neg In(z,H) \vee Owns(J,z)$

10: ~~[illegible]~~ $In(f(H),H)$

13: ∴ $Owns(J, f(H))$ $\theta: \{z/f(H)\}$

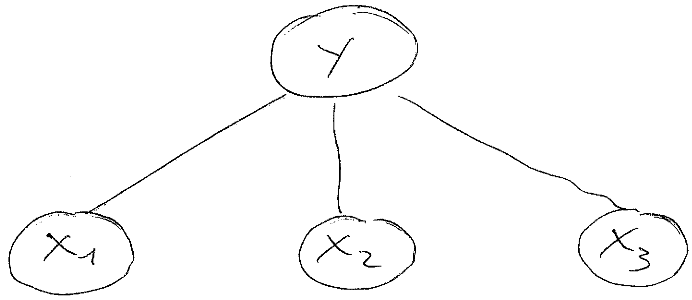
13: $Owns(J, f(H))$

11: $\neg Owns(J, f(H))$

∴ $\emptyset$

EX. 5 30 min

a)



	y = 0	y = 1
$x_1 = 1$	7	7
$x_2 = 1$	1	1
$x_3 = 1$	3	5

$h_{\bar{\theta}}$: $\bar{\theta} = (\theta, \theta_{10}, \theta_{11}, \theta_{20}, \theta_{21} -)$

$P(y = 1) = \theta$

Pr	y = 0	y = 1
$x_1 = 1$	θ_{10}	θ_{11}
$x_1 = 0$	$1-\theta_{10}$	$1-\theta_{11}$

Pr	y = 0	y = 1
$x_2 = 1$	θ_{20}	θ_{21}
$x_2 = 0$	$1-\theta_{20}$	$1-\theta_{21}$

Pr	y = 0	y = 1
$x_3 = 1$	θ_{30}	θ_{31}
$x_3 = 0$	$1-\theta_{30}$	$1-\theta_{31}$

$$P(h_{\bar{\theta}} | \bar{d}) = \alpha \, P(\bar{d} | h_{\bar{\theta}}) \, P(h_{\bar{\theta}})$$

product rule

$$P(\bar{d} | h_{\bar{\theta}}) = \prod_{j=1}^{18} P(\bar{d}_j | h_{\bar{\theta}}) = \prod_j P(\bar{x}_j, y_j | h_{\bar{\theta}}) = \prod_j P(\bar{x}_j | y_j, h_{\bar{\theta}}) \, P(y_j \, h_{\bar{\theta}}) =$$

cond. indep.

$$= \prod_{j=1}^{18} \prod_{i=1}^{3} P(x_{ij} | y_j, h_{\bar{\theta}}) \cdot P(y_j \, h_{\bar{\theta}}) =$$

$$= \theta_{10}^{7} (1-\theta_{10})^{8-7} \cdot \theta_{11}^{7} (1-\theta_{11})^{8-7} \cdot \theta_{20}^{1} (1-\theta_{20})^{8-1} \cdot \theta_{21}^{1} \cdot (1-\theta_{21})^{8-1} \cdot \theta_{30}^{3} (1-\theta_{30})^{8-3} \cdot \theta_{31}^{5} (1-\theta_{31})^{8-5} \, \theta^{8} (1-\theta)$$

$$L(\theta) = \log(P(\bar{d} | h_{\bar{\theta}})) = 7 \lg \theta_{10} + \lg(1-\theta_{10}) + 7 \lg \theta_{11} + \lg(1-\theta_{11}) + \ldots$$

$$\frac{dL}{d\theta} = \frac{8}{\theta} - \frac{8}{1-\theta} = \frac{8 - 8\theta - 8\theta}{\theta(1-\theta)} = 0 \Rightarrow \theta = \frac{1}{2}$$

$$\frac{dL}{d\theta_{10}} = \frac{7}{\theta_{10}} - \frac{1}{1-\theta_{10}} = \frac{7 - 7\theta_{10} - \theta_{10}}{\theta_{10}(1-\theta_{10})} = 0 \Rightarrow \theta_{10} = \frac{7}{8}$$

$$\frac{dL}{d\theta_{21}} = \frac{1}{\theta_{21}} - \frac{7}{1-\theta_{21}} = \frac{1 - \theta_{21} - 7\theta_{21}}{\theta_{21}(1-\theta_{21})} = 0 \Rightarrow \theta_{21} = \frac{1}{8}$$

$$\frac{dL}{d\theta_{30}} = \frac{3}{\theta_{30}} - \frac{5}{1-\theta_{30}} = \frac{3 - 3\theta_{30} - 5\theta_{30}}{\theta_{30}(1-\theta_{30})} = 0 \Rightarrow \theta_{30} = \frac{3}{8}$$

$$1 - \theta_{30} = 1 - \frac{3}{8} = \frac{5}{8}$$

$$\theta_{11} = \frac{7}{8}$$

$$\theta_{31} = \frac{5}{8}$$

$$\theta_{20} = \frac{1}{8}$$

REMOVED

b) $$P(Y \mid \bar{x}, h_{\bar{\sigma}}) \overset{\text{Bayes}}{=} \frac{P(\bar{x} \mid Y, h_{\sigma})\, P(Y, h_{\sigma})}{P(\bar{x})} \overset{\text{cond. indep.}}{=} \frac{\prod_{i=1}^{3} P(x_i \mid Y, h_{\bar{\sigma}})\, P(Y, h_{\bar{\sigma}})}{P(\bar{x})} =$$

$$= \alpha \prod_{i=1}^{3} P(x_i \mid Y, h_{\bar{\sigma}})\, P(Y, h_{\bar{\sigma}}) =$$

$$= \alpha \langle P(x_1=0 \mid y=1) P(x_2=1 \mid y=1) P(x_3=1 \mid y=1)\, P(y=1) ; P(x_1=0 \mid y=0) P(x_2=1 \mid y=0)\, P(x_3=1 \mid y=0)\, P(y=0) \rangle$$

$$= \alpha \langle (1-\sigma_{11})(\sigma_{21})\sigma_{31}\,\sigma ; (1-\sigma_{10})\,\sigma_{20}\,\sigma_{30}\,(1-\sigma) \rangle =$$

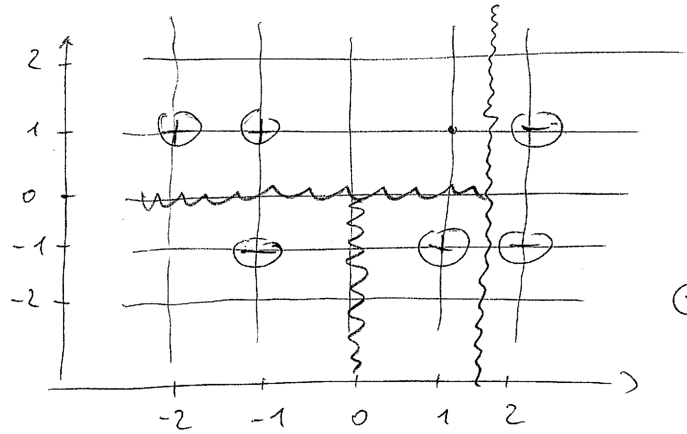
$$= \alpha \langle \frac{1}{8}\,\frac{1}{8}\,\frac{5}{8}\cdot\frac{1}{2} ; \frac{1}{8}\cdot\frac{1}{8}\cdot\frac{3}{8}\cdot\frac{1}{2} \rangle =$$

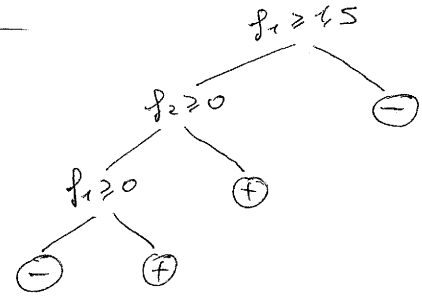
$$= \alpha \langle \; --- \; ; \; -- \; \rangle$$

$y=1$ is more likely than $y=0$ hence we conclude that $y=1$ is the most probable prediction.

EX. 6 — 30 min

a)





b) We calculate the inf. gain at each split point of f_1 and f_2.

$$I(p(v_1), p(v_2)) = \sum_{i=1}^{2} - p(v_i) \lg_2 p(v_i)$$ ← entropy

$$gain(att(s_1)) = I\left(\frac{p}{p+n}; \frac{n}{p+n}\right) - Remainder(attr(s_1))$$

→ before split → after split

$attr(s_1)$ split attribute in s_1

$$Remainder(attr(s_1)) = \sum_{i=1}^{2} \frac{p_i+n_i}{p+n} \; I\left(\frac{p_i}{p_i+n_i}; \frac{n_i}{p_i+n_i}\right)$$

↳ prob. of choosing that branch ↳ inf. in the branch.

$$gain(f_1(+1,5)) = I\left(\frac{3}{6}; \frac{3}{6}\right) - \left(\frac{2}{6} I(0,1) + \frac{4}{6} I\left(\frac{3}{4}; \frac{1}{4}\right)\right) =$$

$$= +\frac{1}{2}\lg\frac{1}{2} - \frac{1}{2}\lg\frac{1}{2} - \left(\frac{1}{3}\left(-0\lg 0 - 1\lg 1\right) + \frac{2}{3}\left(-\frac{3}{4}\lg\frac{3}{4} - \frac{1}{4}\lg\frac{1}{4}\right)\right) =$$

$$= +1 - \frac{2}{3}(0,31 + 0,5) = 0,46$$

$$gain(f_1(0)) = I\left(\frac{3}{6}; \frac{3}{6}\right) - \left(\frac{3}{6} I\left(\frac{1}{3}; \frac{2}{3}\right) + \frac{3}{6} I\left(\frac{2}{3}; \frac{1}{3}\right)\right) = 1 - \left(\frac{1}{2}(0,53+0,39) + \frac{1}{2}(0,53+0,39)\right) =$$

$$= 1 - (0,92) = 0,08$$

$$gain(f_1(-1,5)) = 1 - \left(\frac{1}{6} I(1,0) + \frac{5}{6} I\left(\frac{2}{5}; \frac{3}{5}\right)\right) = 1 - \left(\frac{5}{6}(0,53 + 0,44)\right) = 0,191$$

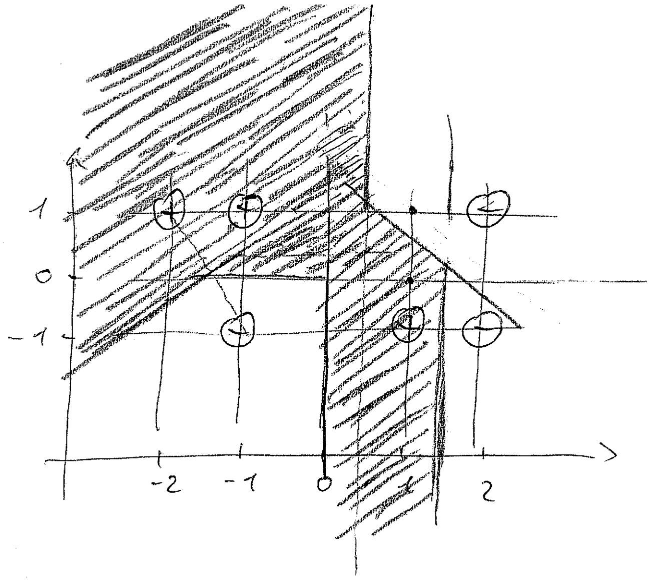
$$gain(f_2(0)) = 1 - \left(\frac{1}{2} I\left(\frac{2}{3}; \frac{1}{3}\right) + \frac{1}{2} I\left(\frac{1}{3}; \frac{2}{3}\right)\right) = 1 - 0,92 = 0,08$$

Hence we select $f_1(+1,5)$ because it maximizes the gain.

c) Prediction for 1,1 is (+)

Nearest Neigh.

a)



b) prediction in (1,1) → ⊖

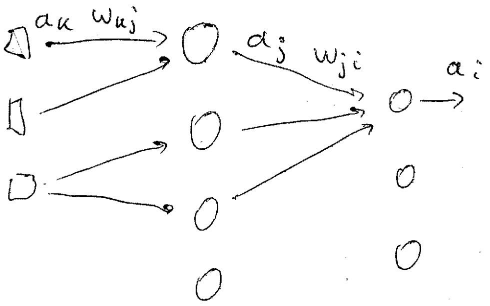
c) prediction by 3-NN → ⊖ for majority rule

d) by crossvalidation

Perceptron

a) The algorithm would not converge because the points are not line separable

Ex. 7 10 min



activation function

$g(x) = cx + d$

$$a_i = g\left(\sum_j a_j w_{ji}\right) = c \sum_j a_j w_{ji} + d = c \sum_j w_{ji}\, g\left(\sum_k w_{kj} a_k\right) + d =$$

$$= c \sum_j w_{ji} \left(c \sum_k w_{kj} a_k + d\right) + d = c^2 \sum_j w_{ji} \sum_k w_{kj} a_k + c \sum_j w_{ji} d + d =$$

$$= c^2 \sum_k \sum_j a_k w_{ji} w_{kj} + \underbrace{\left(c \sum_j w_{ji} + 1\right) d}_{e} =$$

$$= f \sum_k w_{ki} a_k + e$$

where $\sum_k w_{ki} = \sum_k \sum_j w_{ji} w_{kj}$

Hence we could substitute with a single layer network and weights opportunely chosen.