1. $0 <= P(a) <= 1$
   p(true) = 1
   p(false) = 0
   p(a or b) = p(a) + p(b) – p(a and b)

2. p(m | l) = (P (l | m) * p(m)) / p(l)
   p(l | m) = (p(m | l) p (l)) / p(m)
   = .99 * 1 /(10 to the 7) / 1(30)
   = $.297 \times 10^{-6}$

3. Map-maximum a posteriori an approximation of Bayesian prediction where predictions are based on most probable hypothesis. That is an hi that maximizes p(hi | data) * p(hi)

4.

5. p(hot and sunny) = .4 + .1 = .5

6. x1 xor x2 xor x3
   make truth table, and graph the box that is not separable by a plane

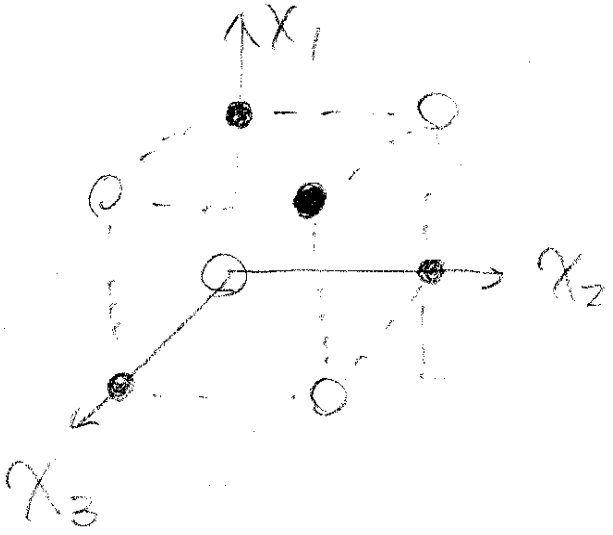
7. w0 = (n - .5)
   wi = 1.0, $1 <= i <= n$
   use sign function
   h(x) = sgn(Σwixi from I = 0 to n - (n - .5)) h(x) = output

8. $delta_j = g'(in_j)\ sum(w_{ji}\ delta_i)$
   $W_{k,j} <- W_{k,j} + alpha \times a_k \times delta_j$
   back propagation is process of adjusting the weights of the hidden layers based on how much they contributed to the final output error.

9. Mercer's theorem states any kernel of a positive definite linear operator corresponds to an inner product in some feature space. A kernel function is a function of the form K(xi, xj) := F(x_i)\dot F(x_j)

(6) X_1 XOR X_2 XOR X_3

X_1	X_2	X_3	X_1 XOR X_2 XOR X_3
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1



No plane can separate group of On's and Off's

(3) MAP - maximum a posteriori

- an approximation of Bayesian prediction where predictions are based on the most probable hypothesis that is on h_i that maximizes $P(d|h_i) P(h_i)$

so

$h_{map} = h_i$ s.t. $P(d|h_i) P(h_i)$ is largest

(4) Ockham's Razor - The simplest solution is the best

using h_{map} - maximizing $P(d|h_i)P(h_i)$ is equivalent to

minimizing $\frac{1}{P(d|h_i)P(h_i)}$

which is equivalent to minimizing

$$\log_2\left(\frac{1}{P(d|h_i)P(h_i)}\right) = -\log_2 P(d|h_i)P(h_i)$$

$$= -\log_2 P(d|h_i) - \log_2 P(h_i)$$

= # of bits needed to describe data d with hypothesis h_i + # of bits to describe h_i

so MAP is examining via choosing shortest string to explain data