I worked with:

- Jens Kristoffersen
- Dherick Devakumar
- Olav Nomeland
- Sanket Behera

I certify that all solutions are entirely in my own words and that I have not looked at another student's solutions. I have given credit to all external sources I consulted.

2.a

Building the heap: $O(k)$ where k is the size of the heap.

Remaining $n-k$ points take $O(\log k)$ time.

Finding the majority label of the k closest neighbours take $O(k)$ time.

Finding n euclidian distances take $n\,O(d)$ time.

Total: $O(k) + (n-k)\,O(\log k) + O(k) + n\,O(d) = O(nd) + O(k) + O(n \log k)$

2.b

Using the problem of placing balls in boxes, the number of monomials is $\binom{d+p}{p}$ and this is the dimension the point lives in.

d is now $\frac{(d+p)!}{(d+p-p)!\,p!} = \frac{(d+p)!}{d!\,p!}$.

New runtime:

$$O\left(n \cdot \frac{(d+p)!}{d!\,p!}\right) + O(k) + O(n \log k)$$

2.c

1D: left and right = 2

2D: all sides and all corners
= 8

dD: There are $2 \cdot d$ adjacent
cells and 2^d corners
$= 2d + 2^d$

You would need to
check each one of these cells so
time complexity is $O(2d + 2^d)$

3.a

I will be 0 surprised since I knew with 100% certainty what the outcome would be.

3.6.

Max surprised, my prior belief was that there was a 0% chance of picking up a white ball.

3.c

when P_B is either 0 or 1 is when the entropy is minimized.

$H_b(0) = -0 \log 0 - 1 \log 1$

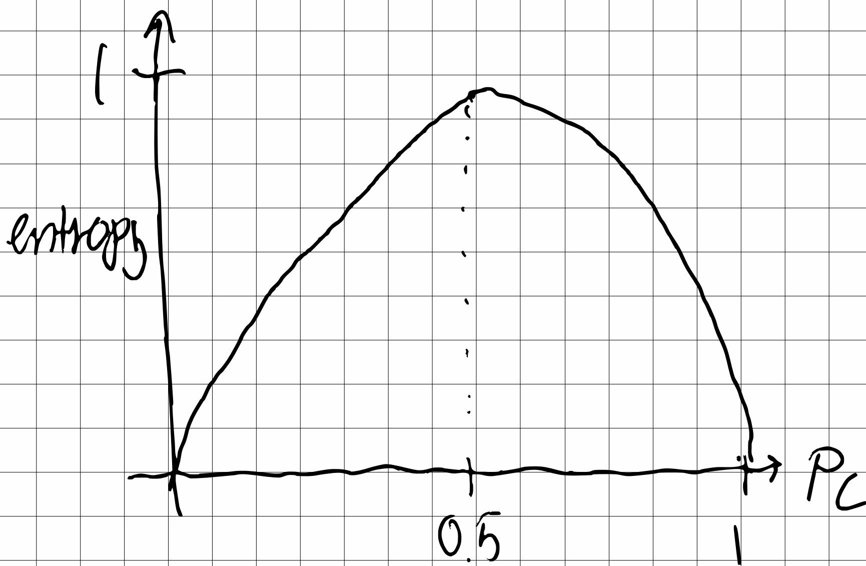
$= \underline{\underline{0}}$

Maximized at $P_B = 0.5$

$Hb(0.5) = -0.5 \log 0.5 - 0.5 \log 0.5$

$= \underline{\underline{1}}$

3. d



strictly concave

3. e.

$$P = P(Y=1 \mid X_{j,v}=1)P(X_{j,v}=1)$$
$$+ P(Y=1 \mid X_{j,v}=0)P(X_{j,v}=0)$$
$$= q_1 P(X_{j,v}=1) + q_2 P(X_{j,v}=0)$$
$$= \lambda q_1 + (1-\lambda) q_2$$ where

λ is the probability that $X_j < v$.

Show that $H(Y) - H(Y \mid X_{j,v})$ is always positive if $q_1 \neq q_2$

$$\to H_b(P(Y=1)) - \sum P(X_{jv}=i) H_b(P(Y=1 \mid X_{j,v}=i))$$
$$> 0$$
$$\to H_b(\lambda q_1 + (1-\lambda)q_2) > \lambda \cdot H_b(q_1) + (1-\lambda)H_b(q_2)$$
$$\to -(\lambda q_1 + (1-\lambda)q_2) \log(\lambda q_1 + (1-\lambda)q_2)$$
$$-(1-\lambda q_1 + (1-\lambda)q_2) \log(1-\lambda q_1 + (1-\lambda)q_2) >$$

$\lambda \cdot (-q_1 \log q_1 - (1-q_1)\log(1-q_1))$
$+ (1-\lambda)\cdot(-q_2 \log q_2 - (1-q_2)$
$\log(1-q_2))$

→ sees that as H_b is strictly concave, multiplying with lamda on the outside of Hb function will lead to a smaller value.

I do not have abetter explanation :/

3.f

Average:

$$E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i]$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n} \mu = \frac{1}{n}\cdot n\mu = \mu$$

Variance:

$$\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(Y_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = \frac{\cancel{n}\sigma^2}{n^{\cancel{2}}} = \frac{\sigma^2}{n}$$

3.9. i

Prob of not being picked:

$$\left(\frac{n-1}{n}\right)^n.$$

Lowest value of this function is at $n=25$ since it is monotonically increasing.

At $n=25$:

$$\left(\frac{25-1}{25}\right)^{25} = 0.3604 > 0.36$$

And at its limit:

$$\lim_{n\to\infty}\left(\frac{n-1}{n}\right)^n = \lim_{n\to\infty}\left(\frac{\cancel{n}\left(1-\frac{1}{n}\right)}{\cancel{n}}\right)^n$$

$$= \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = \lim_{n\to\infty} e^{n\ln\left(1-\frac{1}{n}\right)}$$

$= \frac{1}{e}$ □ $\rightarrow$ P lies in $[0.36, \frac{1}{e}]$

3.g.ii

More trees → more computation

Plot the loss at different nr. of trees in an increasing manner. and pick the number of trees when the loss levels sufficiently out towards its asymptote.

3.h

$$var\left(\frac{1}{n}\sum_{i=1}^{n} z_i\right) = \frac{1}{n^2} var\left[\sum_{i=1}^{n} z_i\right]$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} var(z_i) + 2\sum_{i<j}^{n} cov(z_i, z_j)\right)$$

$$= \frac{1}{n^2}\left(n\sigma^2 + 2\rho\frac{(n-1)n}{2}\right)$$

$$= \frac{\sigma^2}{n} + \frac{\rho(n-1)}{n} \quad \square$$