Answers: Exercise Sheet No. 10

1 OLS: Converting to suitable form

1. $r(\theta) = p - \epsilon r(\theta)\cos(\theta)$
2. $\log(N) = \log(N_0) + (-\lambda t)$
3. $\log(r) = \log(k) + \alpha_1 \log(C_a) + \alpha_2 \log(C_b)$

2 Ordinary Least squares

1. See the lecture slides. Essentially this equation arises from applying first order optimality condition analytically to the problem.

2. (a) $\ln(P) = \beta_1 + \beta_2 T$

   (b)
```
import numpy as np
T = np.array([270.4, 270.6, 272.3, 273.6, 274.1, 275.5,
    276.2, 277.1])
exP = np.array([1.502, 1.556, 1.776, 2.096, 2.281, 2.721,
    3.001, 3.556])
exlnP = np.log(exP)
```

```
X = np.array([np.ones(len(T)), T]).T
Y = exlnP
beta = np.linalg.solve(X.T @ X, X.T @ Y)
```

   (d)
```
N=100
T_lin = np.linspace(270.4, 277.1, N)

log_est = beta[0] + beta[1]*T_lin
est = np.exp(log_est)

import matplotlib.pyplot as plt
plt.plot(T_lin, log_est)

plt.plot(T, exlnP,'o')

plt.xlabel('T')
plt.ylabel('log(P)')
plt.show

plt.plot(T_lin, est)

plt.plot(T, exP,'o')

plt.xlabel('T')
plt.ylabel('P')
plt.show()
```

   (e) In general care must be taken when extrapolating with models fitted to data. This is especially true when a simplified or completely empirical model form is used to estimate parameters. In general, one cannot trust the predictions of this model out of the data range used to estimate the parameters.

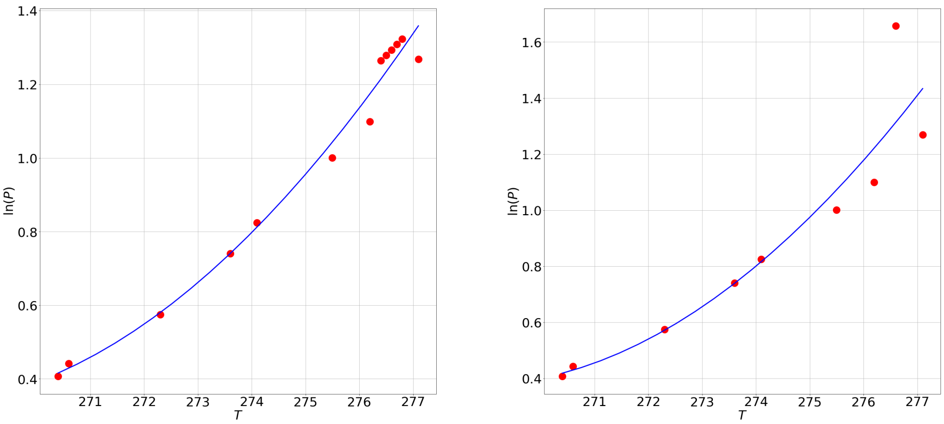

Figure 1: Illustration of the effect of outliers. Left: error $P_\varepsilon = 0.5$ spread over 5 points. Right: error $P_\varepsilon = 0.5$ in a single point.

3.

```
X = np.array([np.ones(len(T)), T, T**(-2)]).T
Y = exlnP
beta2 = np.linalg.solve(X.T @ X, X.T @ Y)

log_pred1 = beta[0] + beta[1]*T
pred1 = np.exp(log_pred1)

log_pred2 = beta2[0] + beta2[1]*T + beta2[2]*T**-2
pred2 = np.exp(log_pred2)
res1 = np.linalg.norm(exP - pred1, 2)
res2 = np.linalg.norm(exP - pred2, 2)

```

(c) The second model. It has to have a lower error because it "contains" the first model, i.e. the data could suggest $\beta_3 = 0$.

(d) No. More complex models in general have a lower error when compared to the fitting data because they are more complex. They may not generalise better to new data.

(e) Use the same code as above with model 2, but use the new T data points.

(f) The plot should look like Figure 1. OLS is more sensitive to outliers (when there is a lot of error in a single point) then when there is distributed error. This is because the squaring in the objective leads to large residuals being much more important than smaller residuals. This is linked to the assumption of normally distributed errors in the MLE formulation, as outliers invalidate this assumption as the tails of the normal distribution are very flat.