1. **Univariate Gaussians**

(a) $x \sim N(1,2) \rightarrow \frac{x-1}{\sqrt{2}} \sim N(0,1)$ $\quad P[x \in (0.5,2)] = P\left[\frac{x-1}{\sqrt{2}} \in \left(-\frac{\sqrt{2}}{4}, \frac{\sqrt{2}}{2}\right)\right] = \phi\left(\frac{\sqrt{2}}{2}\right) - \left(1 - \phi\left(\frac{\sqrt{2}}{4}\right)\right) = \phi\left(\frac{\sqrt{2}}{2}\right) + \phi\left(\frac{\sqrt{2}}{4}\right) - 1 \approx 0.398$

(b) $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

$when\ \sigma\ and\ \mu\ are\ fixed, it\ can\ be\ easily\ conclude\ that\ f(x)\ is\ in\ reverse\ proportion\ with\ (x-\mu)^2, so\ f(x)\ is\ maximized\ when\ x\ is\ equal\ to\ \ \mu\ and\ the\ value\ is\ \frac{1}{\sqrt{2\pi}\sigma}$

(c) $f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$ $\ f(x_1, x_2, \dots \dots, x_n) =$

$\prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left(\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)$

(d) $according\ to\ the\ properties\ of\ mutidimentional\ normal\ distribution, we\ can\ conclude\ that\ (x_1$ $y_1, x_2 + y_2, \dots \dots, x_n + y_n)\ follows\ N\big(x + y; 0, \Sigma_x + \Sigma_y\big)\ and\ (ax_1, ax_2, \dots \dots, ax_n)\ follows\ N(ax; 0, a^2\Sigma_x)$

2. **Logistic regression: basic intuition with one-dimensional data**

(1) Conditional(1) makes 1 classification errors

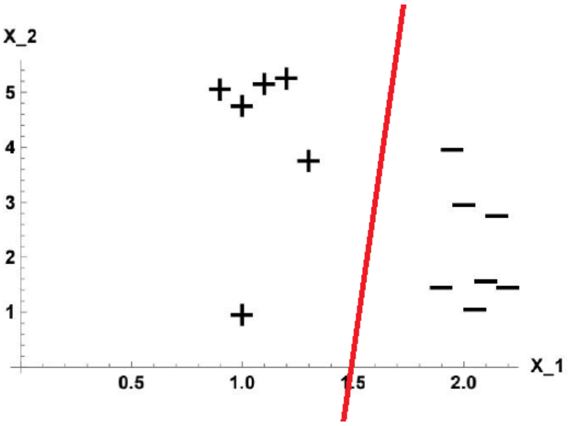
Conditional(2) makes 1 classification errors

(2) the first hypotheses(constant) assigns a higher likelihood to the data according to the joint distribution law,given that each sample point is independent,the joint conditional probability of $P(y1|x1, y2|x2, y3|x3)$ of the first hypotheses is $\frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{4}{9}$ compared with $1 \times 1 \times 0 = 0$ of the second one.

(3) When conditional(1) is the result of the learning,the expectation of $L(g,a)$ on condition that the output is 1 is $E[L(g,a)] = P(y=1|x=-1) \times L(1,1) + P(y=0|x=-1) \times L(0,1) = \frac{20}{3}$ while it is $E[L(g,a)] = P(y=1|x=-1) \times L(1,0) + P(y=0|x=-1) \times L(0,0) = \frac{1}{3}$ when the output is 0,so it finally outputs 0.

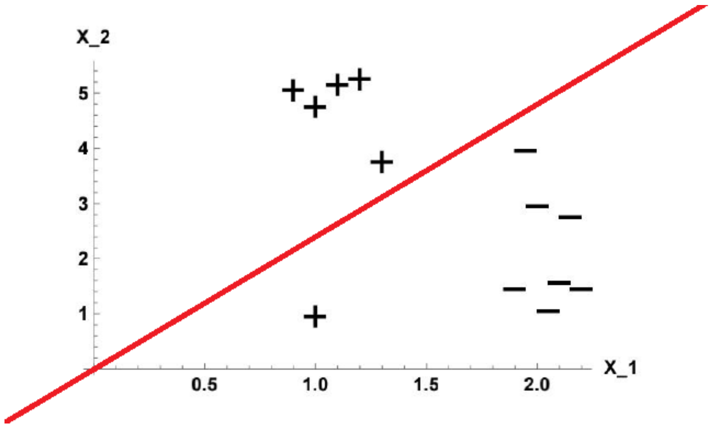
When conditional(2) is the result of the learning,the expectation of $L(g,a)$ on condition that the output is 1 is $E[L(g,a)] = P(y=1|x=-1) \times L(1,1) + P(y=0|x=-1) \times L(0,1) = 10$ while it is $E[L(g,a)] = P(y=1|x=-1) \times L(1,0) + P(y=0|x=-1) \times L(0,0) = 0$ when the output is 0,so it finally outputs 0 as well.

3. **Regularized Logistic Regression**

(a) 

(b) It is not unique because the data set is linearly seperable and there are infinite amount of boundries that can separate the data without classification error.

(c) It makes no classification errors on the training set.

(d) 

(e) Now it makes one classification error on the training set.