**MLISP: Machine Learning in Signal Processing**

# Solutions to problem set 3

*Prof. V. I. Morgenshtern*

*Solver: S. Subic and V. I. Morgenshtern*

## Problem 1: Expected maximum of iid Gaussian random variables

To solve the problem, we will use a standard bounding technique based on the moment generating function (MGF). Consider the MGF of $y$:

$$M_y(t) = \mathbb{E}[e^{ty}] = \mathbb{E}[e^{t \max_i z_i}]. \tag{1}$$

Because the exponential function always takes nonnegative values, we have

$$e^{t \max_i z_i} \leq \sum_i e^{t z_i}. \tag{2}$$

The above is true because among the terms on the left hand side, there is the one on the right hand side and all other terms are positive.

This gives us the following upper bound on the MGF:

$$M_y(t) \leq \mathbb{E}[\sum_i e^{t z_i}] = \sum_i \mathbb{E}[e^{t z_i}]. \tag{3}$$

Since $z_i$ is Gaussian, it's easy to compute it's MGF:

$$\begin{aligned}
\mathbb{E}[e^{t z_i}] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{t z_i} e^{-\frac{z_i^2}{2\sigma^2}} dz_i \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{t z_i - \frac{z_i^2}{2\sigma^2}} dz_i \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\frac{z_i}{\sigma} - t\sigma)^2 + \frac{1}{2} t^2 \sigma^2} dz_i \\
&= e^{\frac{1}{2} t^2 \sigma^2} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\frac{z_i}{\sigma} - t\sigma)^2} dz_i \\
&= e^{\frac{1}{2} t^2 \sigma^2}.
\end{aligned}$$

Combining this with (3) we get:

$$M_y(t) \leq \sum_i e^{\frac{1}{2} t^2 \sigma^2} = N e^{\frac{1}{2} t^2 \sigma^2}. \tag{4}$$

On the other hand, because the exponent is a convex function, by Jensen's inequality:

$$M_y(t) = \mathbb{E}[e^{ty}] \geq e^{t \mathbb{E}[y]}. \tag{5}$$

Combining this with (4) we see

$$e^{t\mathbb{E}[y]} \leq N e^{\frac{1}{2}t^2\sigma^2} \tag{6}$$

Taking the log on both sides gives

$$t\mathbb{E}[y] \leq \log N + \frac{1}{2}t^2\sigma^2 \tag{7}$$

so that

$$\mathbb{E}[y] \leq \frac{\log N + \frac{1}{2}t^2\sigma^2}{t}. \tag{8}$$

This inequality is true for every $t$. To obtain the tightest bound, we can minimize the right hand side across all $t$. Taking the derivative w.r.t. $t$ and setting it to zero be find:

$$-\frac{\log N}{t^2} + \frac{1}{2}\sigma^2 = 0 \implies \frac{\log N}{t^2} = \frac{1}{2}\sigma^2 \implies t^2 = \frac{2\log N}{\sigma^2} \implies t = \frac{\sqrt{2\log N}}{\sigma}.$$

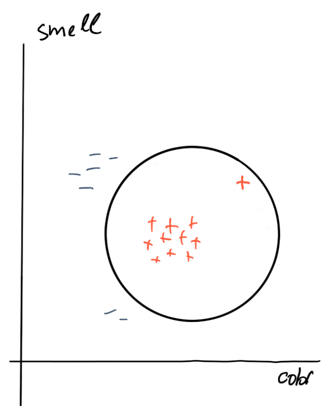
Finally,

$$\mathbb{E}[y] \leq \frac{\log N + \frac{1}{2}\frac{2\log N}{\sigma^2}\sigma^2}{\frac{\sqrt{2\log N}}{\sigma}} = \frac{2\log N}{\sqrt{2\log N}}\sigma = \sigma\sqrt{2\log N}$$

as required.

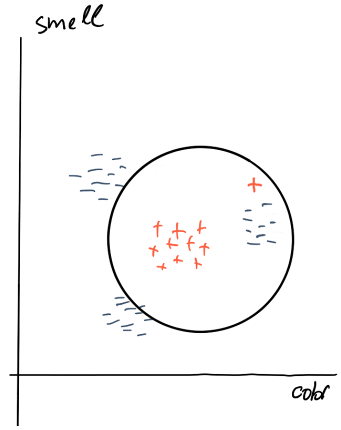
**Problem 6: Debugging machine learning algorithms (exam practice)**

1. One explanation might be that our sample size was too small and the training set looked like this:




Note one positive outlier next to the found decision boundary. This outlier titled the ideal decision boundary to the right.

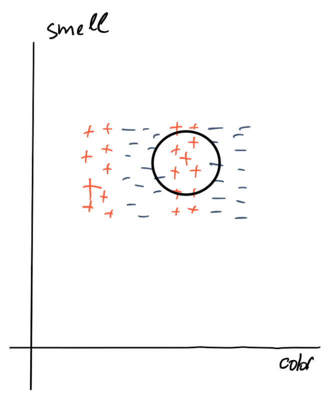
The true distribution looks like this:




Many negative examples on the right cross the decision boundary. Because the training set was too small we did not see the negative examples from the cluster on the right by chance.

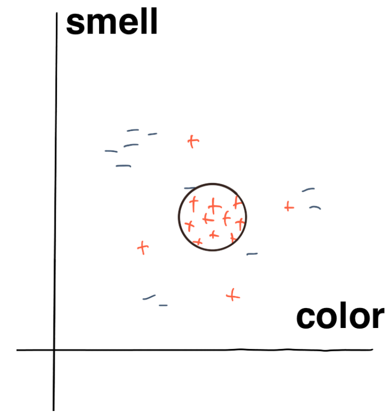
One solution might be to obtain more training data. Another solution might be to make extra modelling assumptions and implement those via regularization. For example we might assume that the good melons are tightly clustered around some optimal parameters. Therefore, the decision circle should be as small as possible. We might put regularization on the radius of the circle and penalize circles of large radius.

2. One explanation might be that the training data looks like this, and therefore cannot be separated by a circle.




The solution is to change the model. For example, the nearest neighbors will work well for this data.

3. By looking at the data, we could decide to penalize the size of the circle: find the smallest circle that is still consistent with the data. This can be achieved by setting $\Omega(r, w_1, w_2) = r$ and for some positive $\lambda$, the solution we get looks like this:




Changing $\lambda$ makes the circle bigger or smaller.

**Problem 7: Haar wavelets (exam practice)**

1. $V_0$ is one dimensional space (when restricted to the interval $[0,1]$). Its only basis element is

$$\phi_{0,0}(x) = 1, x \in [0,1].$$

$W_0$ is one dimensional space (when restricted to the interval $[0,1]$). Its only basis element is

$$\psi_{0,0}(x) = \begin{cases} 1, & x \in [0, \frac{1}{2}) \\ -1, & x \in [\frac{1}{2}, 1]. \end{cases}$$

$W_1$ is two dimensional space (when restricted to the interval $[0,1]$). Its basis elements are

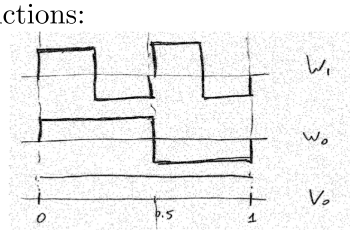
$$\psi_{1,0}(x) = \begin{cases} \sqrt{2}, & x \in [0, \frac{1}{4}) \\ -\sqrt{2}, & x \in [\frac{1}{4}, \frac{1}{2}) \\ 0, & x \in [\frac{1}{2}, 1]. \end{cases}$$

and

$$\psi_{1,1}(x) = \begin{cases} 0, & x \in [0, \frac{1}{2}) \\ \sqrt{2}, & x \in [\frac{1}{2}, \frac{3}{4}) \\ -\sqrt{2}, & x \in [\frac{3}{4}, 1]. \end{cases}$$

Therefore, $\dim(V_0, W_0, W_1) = (1, 1, 2)$.

Here are the plots of these four functions:




2.

$$f = \langle f, \phi_{0,0} \rangle \, \phi_{0,0} + \langle f, \psi_{0,0} \rangle \, \psi_{0,0} + \langle f, \psi_{1,0} \rangle \, \psi_{1,0} + \langle f, \psi_{0,1} \rangle \, \psi_{0,1}.$$

This is true because the functions $\phi_{0,0}, \psi_{0,0}, \psi_{1,0}, \psi_{0,1}$ form an orthonormal set.

3.

$$\langle f, \psi_{1,0} \rangle = \int_0^1 f(x)\psi_{1,0}(x)dx = 0$$
$$\langle f, \psi_{0,1} \rangle = \int_0^1 f(x)\psi_{0,1}(x)dx = 0.$$

4. The function $f$.