NEIS Insight

December 2020

```
each: function(e, t, n) {
    var r, i = 0,
        o = e.length,
        a = M(e);
    if (n) {
        if (a) {
            for (; o > i; i++)
                if (r = t.apply(e[i], n), r === !1) break
        } else
            for (i in e)
                if (r = t.apply(e[i], n), r === !1) break
    } else if (a) {
        for (; o > i; i++)
            if (r = t.call(e[i], i, e[i]), r === !1) break
    } else
        for (i in e)
            if (r = t.call(e[i], i, e[i]), r === !1) break;
    return e
},
trim: b && !b.call("\ufeff\u00a0") ? function(e) {
    return null == e ? "" : b.call(e)
} : function(e) {
    return null == e ? "" : (e + "").replace(C, "")
},
makeArray: function(e, t) {
    var n = t || [];
    return null != e && (M(Object(e)) ? x.merge(n, "string" =
},
inArray: function(e, t, n) {
    var r;
    if (t) {
        if (m) return m.call(t, e, n);
        for (r = t.length, n = n ? 0 > n ? Math.max(0, r + n
            if (n in t && t[n] === e) return n
    }
```

MADS ADD ON

Are you already data minded ?

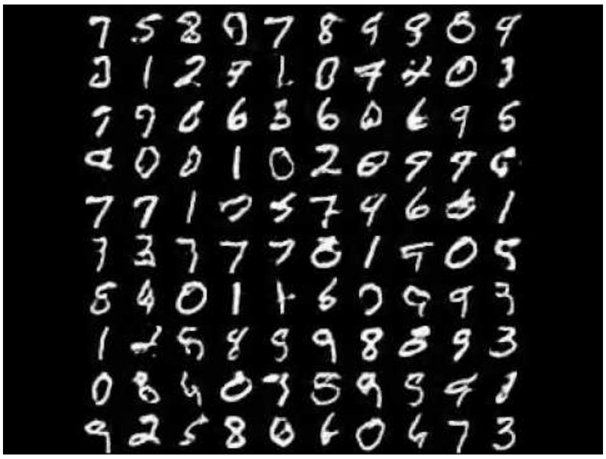
A fun quiz

Just an example of data science which shows, what kind of things can be seen as data and which transformation is necessary so that the machine can learn from these data.

The fun quiz is if you can guess the solution path?

At the same time it is an illustration of a problem which is difficult to solve with classical programming.

THE PROBLEM



Practically you know the issue and know that it is solved.

It is OCR - optical character recognition.

Here you see the principle how it became solvable.

How to recognize handwritten figures ?

First one takes a probe of handwritten figures from 0 to 9 from many different individuals. Then there are scanned in, in order to get digital pictures of these figures.

As everybody knows a figure looks quite different depending on who has written it. Thus your employee concludes to program this might be quite difficult. Hence he wants to feed these new machine learning algorithms with the problem and get it solved that way.

But how to feed machine learning algos with pictures ?

He approaches You as his manager and ask for support.

Can You give him direction ?

WHAT DO YOU THINK SHOULD BE DONE ?

Do not flip the page / go to the next page before having noted your solution to the problem!

Check out

The internet with the abbreviation MNIST

THE PRINCIPLE SOLUTION

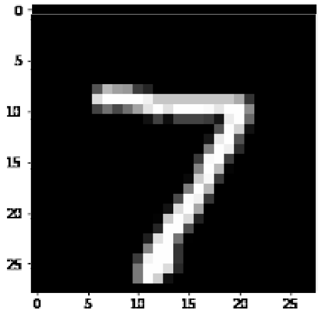
A digital picture comprises of pixels organized into two axes. Each pixel has a value.

For black and white pictures it is a value of a grey tone from black to white. For a color pictures each pixel has 3 values for Red, Green and Blue.

Here the images of handwritten figures have 28 times 28 pixels or 784 pixels. Each pixel has a value of 0 (black) to 255 (white).

In order to work with it the image it is transformed into a vector x with 748 dimensions.

Thus the data we deal with are now vectors and for each vector x_i there is answer y_i with a value from 0 to 9.



Picture 28 x 28 pixels

$\longmapsto$ X e $\mathbb{R}^{784}$

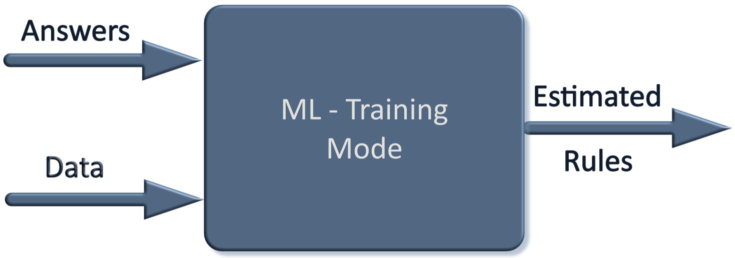
LET THE MACHINE LEARN

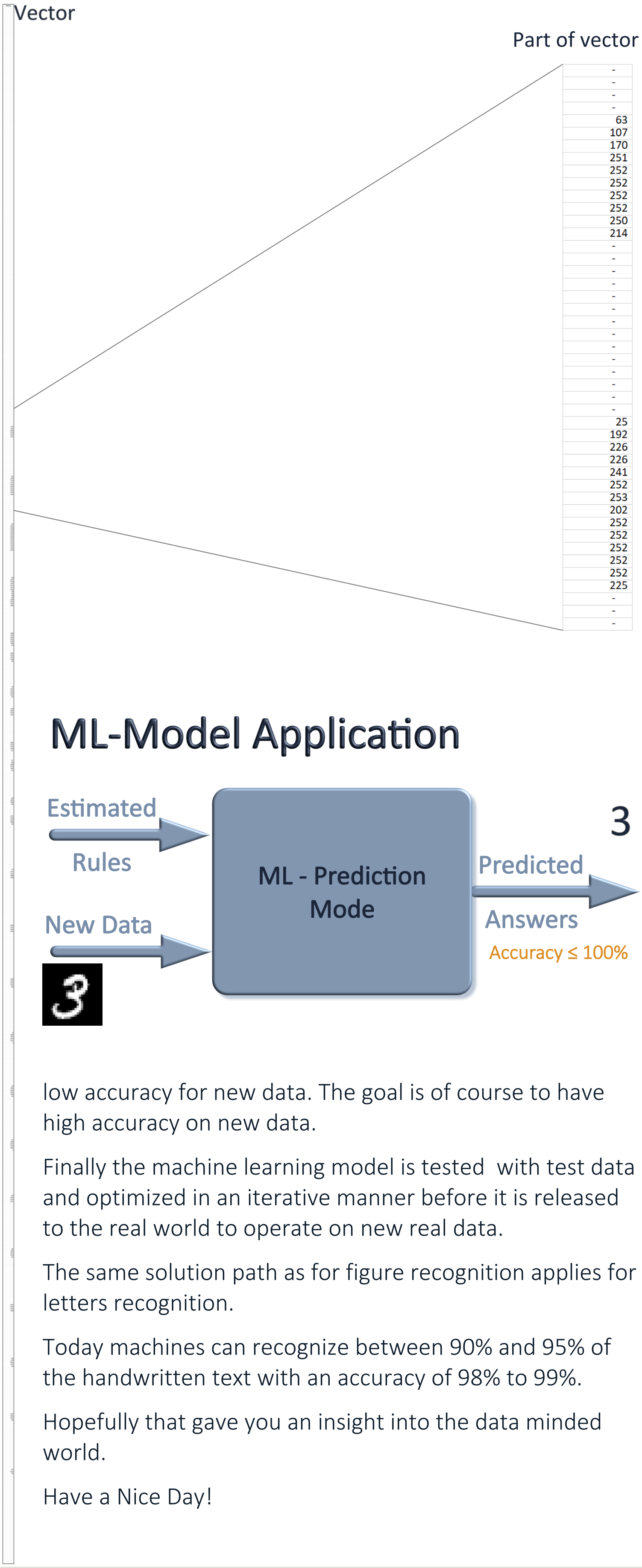
Now that the data in form of vectors are readable to the computer. The learning can be done with machine learning algorithms like a) nearest neighbor, Gaussian generative models, c) neural networks or d) other.

The nearest neighbor e.g. works with Euclidean Distance between the vectors. The others with mathematically more complex constructs.

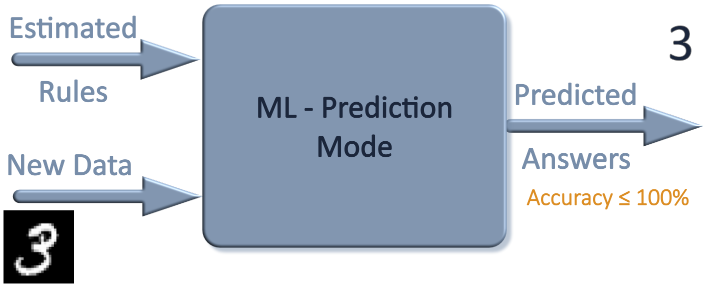
Before learning sometimes the high dimensional data (here 784) are compressed with e.g. Principal component analysis (PCA) to extract only the most important, the most characteristic information. Hence subsequent learning computations are faster.

Machine Learning





ML-Model Application



The result is a trained model which is checked with validation data in order to counterbalance the phenomenon of "overfitting" the model to the training data. Overfitted models have a very high accuracy on the training data, but a low accuracy for new data. The goal is of course to have high accuracy on new data.

Finally the machine learning model is tested with test data and optimized in an iterative manner before it is released to the real world to operate on new real data.

The same solution path as for figure recognition applies for letters recognition.

Today machines can recognize between 90% and 95% of the handwritten text with an accuracy of 98% to 99%.

Hopefully that gave you an insight into the data minded world.

Have a Nice Day!