# Accuracy of COVID tests

True Positive
prediction correct
predicted 1

- The results of 100 Michigan Medicine COVID tests are given below.

confusion matrix

| | Predicted Negative | Predicted Positive |
|---|---|---|
| **Actually Negative** | TN = 90 ✅ | FP = 1 ❌ |
| **Actually Positive** | FN = 8 ❌ | TP = 1 ✅ |

*Michigan Medicine test results*

False Negative
prediction was wrong
predicted 0

- 🤔 **Question:** What is the accuracy of the test?

localhost

## Discussion

$$\text{precision} = \frac{TP}{TP + FP} \qquad \text{recall} = \frac{TP}{TP + FN}$$

- 🤔 When might high **precision** be more important than high recall?
- 🤔 When might high **recall** be more important than high precision?

false positives really bad

e.g. predicting crime, honor code violation

medical tests

?

# Logistic regression

- Logistic **regression** is a linear **classification** technique that builds upon linear regression.

- It models **the probability of belonging to class 1, given a feature vector**:

$$P(y = 1|\vec{x}) = \sigma(\underbrace{w_0 + w_1 x^{(1)} + w_2 x^{(2)} + \ldots + w_d x^{(d)}}_{\text{linear regression model}}) = \sigma\left(\vec{w} \cdot \text{Aug}(\vec{x})\right)$$

nice S shape

for modeling probabilities

why?

$0 < \sigma(t) < 1$ for all $t$

so we can interpret outputs as probabilities

# Attempting to use squared loss

- Our default loss function has always been squared loss, so we could try and use it here.

$$R_{\text{sq}}(\vec{w}) = \frac{1}{n} \sum_{i=1}^{n} \left( y_i - \underbrace{\sigma\left(\vec{w} \cdot \text{Aug}(\vec{x}_i)\right)}_{\text{predicted probabilities!}} \right)^2$$

$0 < p_i < 1$

still just

$$\frac{1}{n} \sum_{i=1}^{n} \left( \text{actual}_i - \text{predicted}_i \right)^2$$

only two possibilities:

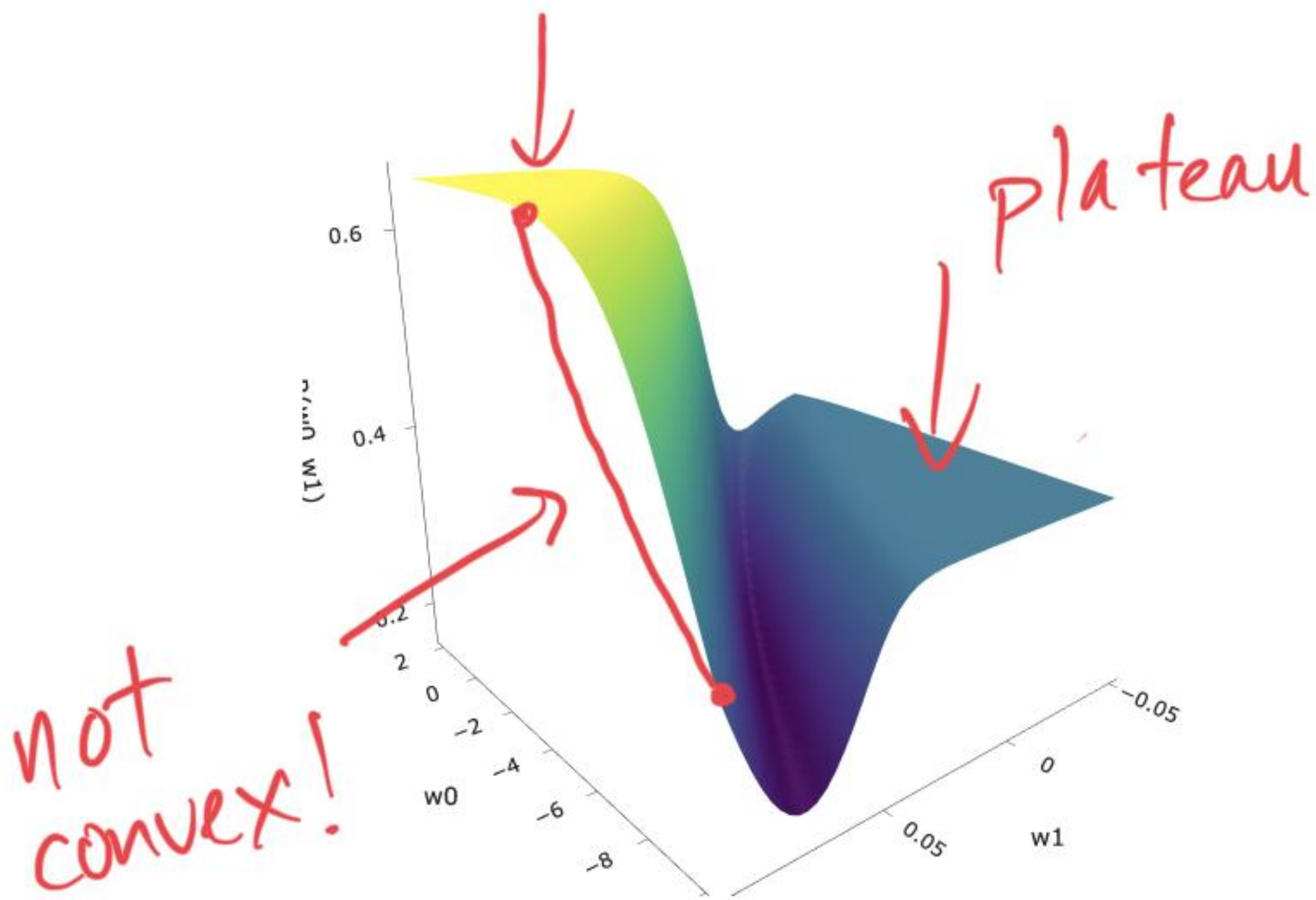
1 , 0

person i actually had diabetes

person i did not have diabetes

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sigma(w_0 + w_1 \underbrace{x_i}_{\text{Glucose}_i})\right)$$

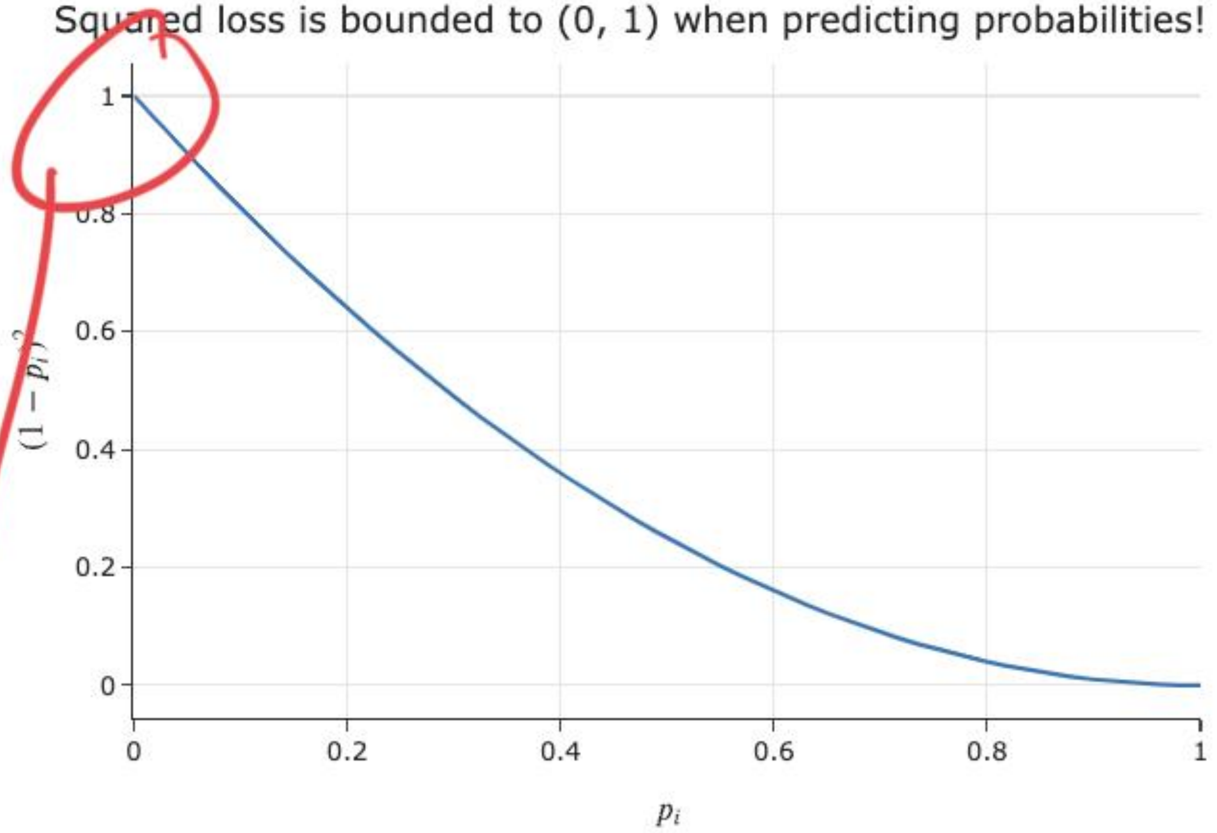
```
In [15]: util.show_logistic_mse_surface(X_train, y_train)
```




- **What do you notice?**

- Suppose $y_i = 1$. Then, the graph of the squared loss of the prediction $p_i$ is below.

In [16]:
```
util.show_squared_loss_individual()
```




$(1-p_i)^2$

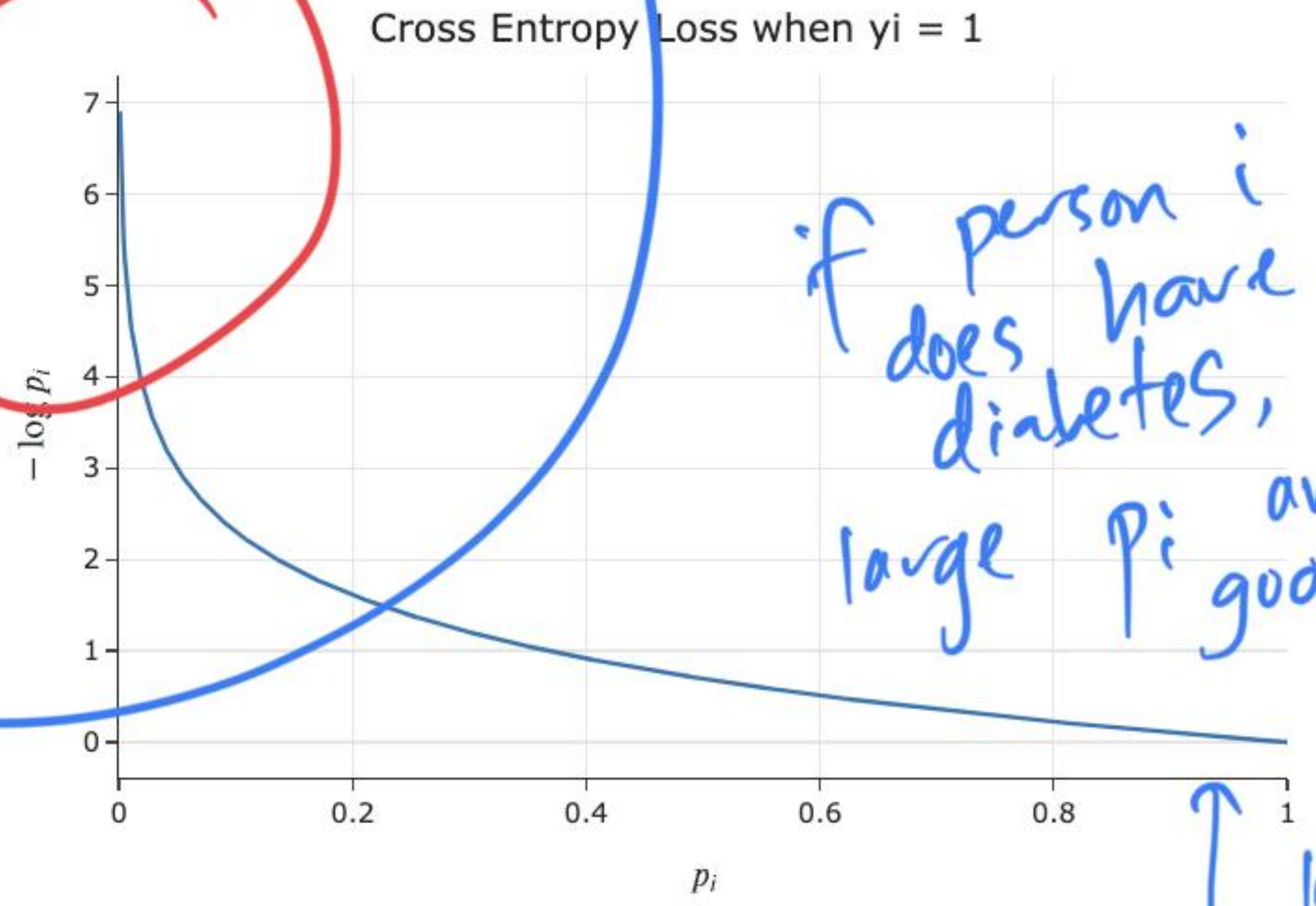
worst/largest possible squared loss when predicting probabilities is $(1-0)^2 = 1$.

$$L_{\text{ce}}(y_i, p_i) = \begin{cases} -\log(p_i) & \text{if } y_i = 1 \\ -\log(1 - p_i) & \text{if } y_i = 0 \end{cases}$$

- Note that in the two cases – $y_i = 1$ and $y_i = 0$ – the cross-entropy loss function resemble squared loss, but is unbounded when the predicted probabilities $p_i$ are far from $y_i$.

bad!

In [17]:
```
util.show_ce_loss_individual_1()
```




if person i does have diabetes, large pi are good!

low loss

as pi → 1,
loss → 0

as pi → 0,
loss → ∞ !

good ✓
strongly penalizes bad predictions

In [ ]:
```
util.show_ce_loss_individual_0()
```

another derivation: maximum likelihood

## A non-piecewise definition of cross-entropy loss

- We can define the cross-entropy loss function piecewise. If $y_i$ is an observed value and $p_i$ is a predicted **probability**, then:

$$L_{\text{ce}}(y_i, p_i) = \begin{cases} -\log(p_i) & \text{if } y_i = 1 \\ -\log(1 - p_i) & \text{if } y_i = 0 \end{cases}$$

equivalent!

- An equivalent formulation of $L_{\text{ce}}$ that isn't piecewise is:

$$L_{\text{ce}}(y_i, p_i) = -\left(y_i \log p_i + (1 - y_i)\log(1 - p_i)\right)$$

$$L_{ce}(1, p_i) = -\log p_i - \underbrace{(1-1)}_{0}\cancel{\log(1-p_i)} = -\log p_i$$

$$L_{ce}(0, p_i) = -\cancel{0 \log p_i} - (1-0)\log(1-p_i) = -\log(1-p_i)$$

?

# Decision boundaries for logistic regression

- In our single feature model that predicts `'Outcome'` given just `'Glucose'`, our predicted probabilities are of the form:

$$P(y = 1|\text{Glucose}) = \sigma\left(w_0^* + w_1^* \cdot \text{Glucose}\right)$$

what is c?

- Suppose we fix a threshold, $T$. Then, our **decision boundary** is of the form:

Glucose = c

$$\sigma^{-1}\left(\sigma\left(w_0^* + w_1^* \cdot \text{Glucose}\right)\right) = \sigma^{-1}(T)$$

$$w_0^* + w_1^* \cdot \text{Glucose} = \sigma^{-1}(T)$$

$$\boxed{\text{Glucose}_T = \frac{\sigma^{-1}(T) - w_0^*}{w_1^*}}$$