

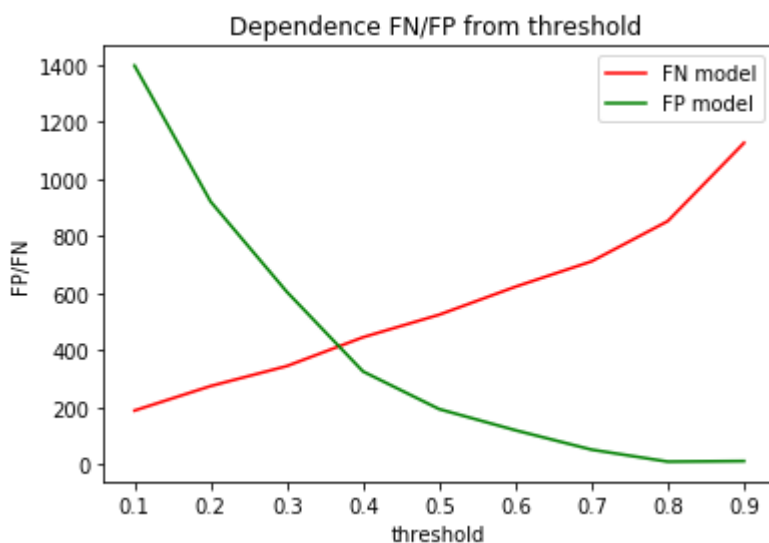
Do thresholds for the highest profit and highest profit per customer coincide or not? Why?

They didn't coincide (max profit with Threshold = 0.2 and max profit per user with Threshold = 0.1) in strategy B. This happens because profit depends on the number of people who we give a discount. The metric "profit per customer" also depends on this parameter because it calculates from profit divides on customers who predicted as not churn.

With changing threshold the profit tends to decrease. I think that reasonable way is to choose 20% threshold because in this situation we will give a discount to less amount of people (than with threshold=10%) so we can save some money and it is more or less a compromise

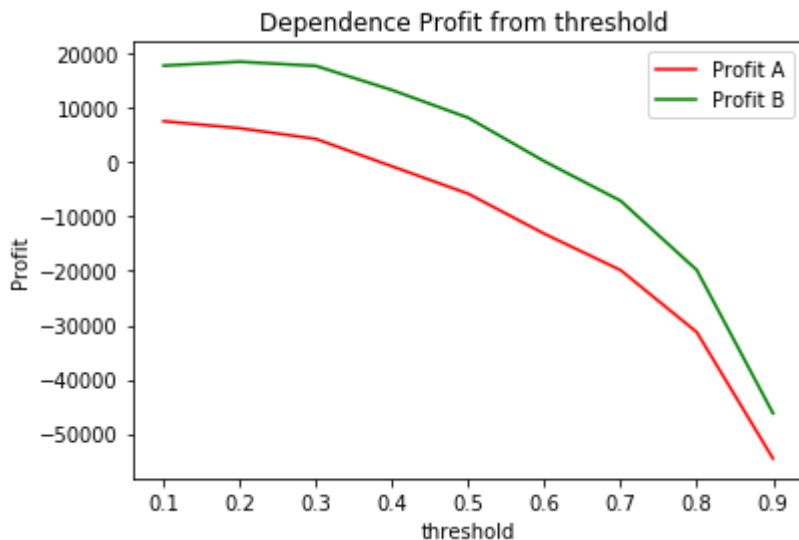
In [62]:

```
1 import matplotlib.pyplot as plt
2 plt.plot([0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9], fn_arr, color = "r", label='FN')
3 plt.xlabel('threshold')
4 plt.plot([0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9], fp_arr, color = "g", label='FP')
5 plt.xlabel('threshold')
6 plt.ylabel('FP/FN')
7 plt.legend()
8 plt.title("Dependence FN/FP from threshold ")
9 plt.show()
```



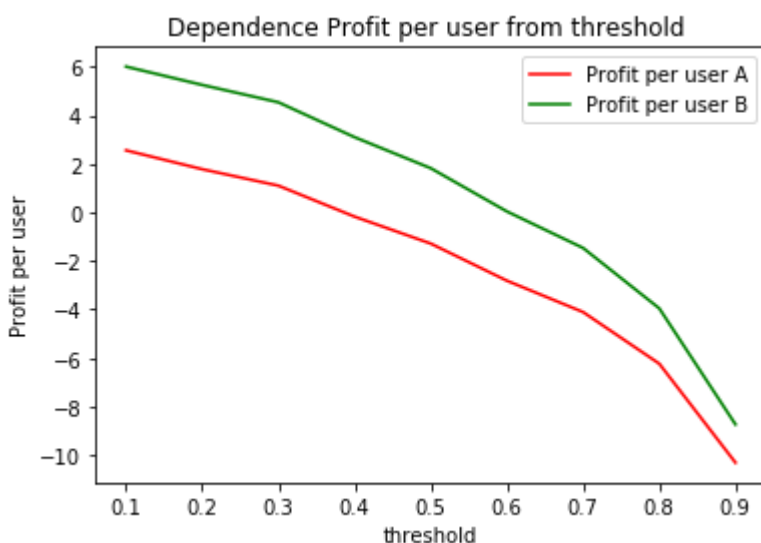
In [63]:

```
1 plt.plot([0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9], profit_a_arr, color = "r", label="Profit A")
2 plt.plot([0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9], profit_b_arr, color = "g", label="Profit B")
3 plt.xlabel('threshold')
4 plt.ylabel('Profit')
5 plt.legend()
6 plt.title("Dependence Profit from threshold ")
7 plt.show()
```



In [64]:

```
1 #profit_a_per_user_arr
2 plt.plot([0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9], profit_a_per_user_arr, color = "r", label="Profit per user A")
3 plt.plot([0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9], profit_b_per_user_arr, color = "g", label="Profit per user B")
4 plt.xlabel('threshold')
5 plt.ylabel('Profit per user')
6 plt.legend()
7 plt.title("Dependence Profit per user from threshold ")
8 plt.show()
```



Which you decide to choose? Under what circumstances (how many clients will you lose in both situations; what should be the decision criteria)?

Actually, this depends on the whole company's economic situation. I guess that in a real situation we will not have such a huge acceptance rate. I like the variant when we optimize the model to identify maximum people who will churn (max TP and min FN) (choose Strategy B). So I will choose the model with the lowest FN, but in a real situation it is necessary to find a balance between models with lowest FN and higher FP. The formula for lost clients will be like:

$FN + (TP * 0.1) = 357 + (1512 * 0.1) = 508.2$ So we will lose about 508 customers.

How hard does your profit per customer decrease for customers for whom you provide a discount, compared to customers for whom you do not provide a discount? Compare this number with the discount.

In [66]:

```
profit_without_discont = tn * p + fp * p - losses
profit_per_user_without_discont = profit_without_discont / collections.Counter(df['Churn'])[0]
print("profit_per_user_without_discont = ", profit_per_user_without_discont)
print("profit_per_user_b = ", profit_per_user_b)
print("p = ", p)
print("30% discount = ", p*0.3)
print("With discount profit per user decreased at = ", round(100 - (profit_per_user_b * 100 / profit_per_user_without_discont), 2))
```

```
profit_per_user_without_discont = 34.7118206062906
profit_per_user_b = 5.265861816009369
p = 62.85776104583815
30% discount = 18.857328313751445
With discount profit per user decreased at = 84.83
```

According to my calculation profit per customer decreases at almost 84%. But providing discount still relevant because we still saving some customers and make more profit from them

In []:

1